

INN: SOBEK et al.

P. Date: 3/9/01

CRFE
SEARCH REQUEST FORM

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

113445

Requestor's
Name:

JANE ZARA

Serial
Number:

09/803,165

Date: 2-6-04

Phone: 2-0765

Art Unit: 1635

2718

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

1181

Please Search Seq ID No: 34

No size limits

Thanks

STAFF USE ONLY

Date completed: 2/10

Searcher: _____

Terminal time: _____

Elapsed time: _____

CPU time: _____

Total time: _____

Number of Searches: _____

Number of Databases: _____

Search Site

STIC

CM-1

Pre-S

Type of Search

N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG

STN

Dialog

APS

Geninfo

SDC

DARC/Questel

Other

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

OM protein - protein search, using sw model1

Run on: February 9, 2004, 11:48:25 ; Search time 28 Seconds
(without alignment)

1169.593 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026

Sequence: 1 MIFDTDXITDGKPIIRFK.....KEDLKYQSSKQVGLDAWLRK 774

Scoring table: BLOSUM62

Gappen: 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3225.5	30.1	774	3 US-08-902-032-2	Sequence 2, Appli
2	3222.5	80.0	774	3 US-09-073-354-1	Sequence 1, Appli
3	3222.5	80.0	774	3 US-08-656-005A-1	Sequence 1, Appli
4	3222.5	80.0	774	3 US-09-073-259-1	Sequence 1, Appli
5	3222.5	80.0	774	3 US-09-363-995-1	Sequence 1, Appli
6	3222.5	80.0	774	3 US-09-418-027-1	Sequence 1, Appli
7	3209.5	79.7	778	2 US-08-906-925-4	Sequence 4, Appli
8	3151.5	78.3	775	1 US-07-966-778-1	Sequence 1, Appli
9	3151	78.3	775	1 US-08-424-921-1	Sequence 1, Appli
10	3151	78.3	775	2 US-08-556-355A-1	Sequence 1, Appli
11	3151	78.3	775	2 US-07-803-27A-1	Sequence 1, Appli
12	3151	78.3	775	2 US-09-244-889A-1	Sequence 1, Appli
13	3151	78.3	776	2 US-08-688-649-17	Sequence 4, Appli
14	3141.5	78.0	773	4 US-09-585-858-15	Sequence 35, Appli
15	3139	78.0	779	5 PCT-US95-152263-12	Sequence 12, Appli
16	3139	78.0	779	5 US-08-375-134-12	Sequence 32, Appli
17	2974	73.9	727	4 US-09-585-958-12	Sequence 8, Appli
18	2502.5	62.2	1022	1 US-08-271-164A-8	Sequence 27, Appli
19	2502.5	62.2	1022	2 US-08-222-115B-27	Sequence 7, Appli
20	2069	51.4	1019	2 US-08-222-715B-26	Sequence 2, Appli
21	2069	51.4	1019	2 US-08-229-884A-2	Sequence 2, Appli
22	1595	39.6	396	1 US-08-907-166-6	Sequence 6, Appli
23	1403	34.8	788	4 US-09-391-340-6	Sequence 6, Appli
24	1403	34.8	788	4 US-08-907-166-4	Sequence 4, Appli
25	1194	29.7	803	2 US-09-391-340-4	Sequence 4, Appli
26	1194	29.7	803	4 US-08-062-368-4	Sequence 4, Appli
27	1191	29.6	803	1 US-08-062-368-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-902-632-2
; Sequence 2, Application US/08902632
; Patent No. 6008025
; GENERAL INFORMATION:
; APPLICANT: KOMATSUBARA, Hideyuki
; APPLICANT: KITABAYASHI, Masao
; APPLICANT: KANIMURA, Hideki
; APPLICANT: KAWAKAMI, Bunsei
; APPLICANT: KAWAMURA, Yoshihisa
; APPLICANT: TAKAGI, Masahiro
; APPLICANT: IMANAKA, Tadayuki
; TITLE OF INVENTION: Modified Thermostable DNA Polymerase,
; and DNA Polymerase Composition for Nucleic Acid
; TITLE OF INVENTION: Amplification
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1 Broadway
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.50 inch floppy disk, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902/32
; FILING DATE: Concurrent Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 198911/956
; FILING DATE: 29-JUL-1996
; APPLICATION NUMBER: JP 200446/96
; FILING DATE: Concurrent Herewith
; APPLICATION NUMBER: 30-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GREASON, Edward W.
; REGISTRATION NUMBER: 18,918
; REFERENCE/DOCKET NUMBER: 2418/7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-425-7200
; TELEFAX: 212-425-5288
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-902-632-2

Query Match Score 3225.5; DB 3; Length 774;

Best Local Similarity 77.3%; Pred. No. 1.2e-24; Matches 599; Conservative 82; Mismatches 89; Indels 3; Gaps 2;

Qy 1 MIFDTDITDGGPKIIRIFKENGKPKIELDPHFPQIYALLKDSADEIAALKGERHG 60

Db 1 MILDTDITDGGPKVIRIFKENGKPKIELDPHFPQIYALLKDSADEIAEVKKTAAERHG 60

Qy 61 KIVRVIIDAVKVYKQKFLDRLVWKKLFLFPHDPDVPAIRGKIREHPAVIDYDIPFARY 120

Db 61 TIVTVTKRVEVKYQKFLGPRVWVKLYFTHFCDVPAIRDKIREHGAVIDYDIPFARY 120

Qy 121 LIDKGJLIPMEGDEELKLMADFLETFYHGDEFGEKGIIMSYADEFEARVITWNIDPY 180

Db 121 LIDKGJLIPMEGDEELKLMADFLETFYHGDEFGEAEGPILMSYADEGARVITWNIDPY 180

Qy 181 VDVVSNSPREM1KRFVQ1VREKDPVDTIYTNGDNFDLPLYKRAEKLGVTLLGRDKEHP 240

Db 181 VDVVSNSPREM1KRFVQ1VREKDPVDTIYTNGDNFDLPLYKRAEKLGVTLLGRDKEHP 240

Qy 241 PKTHRMGDSFAVBIKGTHFDPYVTRTINLPTYLEAVYBAVIGKTKSKLGAEELAAI 300

Db 239 PKIQRMGDRFAVEVKGTHFDPYVTRTINLPTYLEAVYBAVQSPKRVAYEITPTA 298

Qy 301 WETEESMKKLAQYSMEDARATYELGKPEPMEAEELAKLIGOSWVDSRSSSTGNLYEWYL 360

Db 299 WETEESMKKLAQYSMEDARATYELGKPEPMEAEELAKLIGOSWVDSRSSSTGNLYEWYL 358

Qy 361 RAYERNEELAPNPKDECEYLARR-RQSTEGGYKPEPGLWNVNIVLFRSYPSTTIN 420

Db 359 RAYERNEELAPNPKDECEYLARR-RQSTEGGYKPEPGLWNVNIVLFRSYPSTTIN 417

Qy 421 VSDDTLEPEGKNDYAP1VGYKFCRDPFGIPSIGELITMRQETKRRKATIDIEKK 480

Db 418 VSDDTLEPEGKNDYAP1VGYKFCRDPFGIPSIGELITMRQETKRRKATIDIEKK 477

Qy 481 MLDYQRRAVKLHANSYTGYMGYPKARYSKCAGESTAWGHYIETIKEEKGFKVL 540

Db 478 LILYQRRAKILLANSYKQYGYTARARYCKEAESVTAWGHYIETIKEEKGFKVY 537

Qy 541 YADTDGFATATIPGKPKETIKKAKAKFVKYINGKLPGLELEYEGFYLRGFVAKKRYAVI 600

Db 538 YSDPDGFATIPGKPKETIKKAKAKFVKYINGKLPGLELEYEGFYLRGFVAKKRYAVI 597

Qy 601 DERTGTTTGCLEVRDQSRERAKTQVYKLEALKEDSVKAVETVQDVYELAKYQVPL 660

Db 598 DEEKSITTRGLELVRDQSRERAKTQVYKLEALKEDSVKAVIVTEKLSKVBPV 657

Qy 661 EKVLYTHEQITKOLSEYKAIGPHYIAIAKRLAAGKIKVRPGTISIYIVLGRSGSKISDRVILL 720

Db 658 EKVLYTHEQITKOLSEYKAIGPHYIAIAKRLAAGKIKVRPGTISIYIVLGRSGSKISDRVILL 717

Qy 721 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 773

Db 718 DEFPPTKHYDAEYYIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 770

Qy 731 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 771

Db 722 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 772

Qy 732 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 773

Db 723 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 774

Qy 733 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 775

Db 724 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 776

Qy 734 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 777

Db 725 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 778

Qy 735 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 779

Db 726 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 780

Qy 736 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 781

Db 727 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 782

Qy 737 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 783

Db 728 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 784

Qy 738 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 785

Db 729 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 786

Qy 739 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 787

Db 730 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 788

Qy 740 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 789

Db 731 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 790

Qy 742 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 791

TITLE OF INVENTION: Amplifying Nucleic Acids

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon & Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 6.1 Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073,354

FILING DATE: Concurrent Herewith

CLASSIFICATION: 0506

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/656,005

FILING DATE: 24 MAY 1996

APPLICATION NUMBER: JP 134036/95

FILING DATE: 31 MAY 1995

ATTORNEY/AGENT INFORMATION:

NAME: Oleneneti, Judith L.

REGISTRATION NUMBER: 39,048

REFERENCE/DOCKET NUMBER: 2418/9

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-429-1776

TELEFAX: 202-429-0796

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 774 amino acids

REFERENCE/DOCKET NUMBER: 2418/9

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-429-1776

TELEFAX: 202-429-0796

Best Local Similarity 77.4%; Prod. No. 2e-242;

Matches 598; Mismatches 83; Indels 3; Gaps 2;

Qy 1 MIFDTDITDGGPKIIRIFKENGKPKIELDPHFPQIYALLKDSADEIAALKGERHG 60

Db 1 MILDTDITDGGPKVIRIFKENGKPKIELDPHFPQIYALLKDSADEIAEVKKTAAERHG 60

Qy 61 KIVRVIIDAVKVYKQKFLDRLVWKKLFLFPHDPDVPAIRGKIREHPAVIDYDIPFARY 120

Db 61 TIVTVTKRVEVKYQKFLGPRVWVKLYFTHFCDVPAIRDKIREHGAVIDYDIPFARY 120

Qy 121 LIDKGJLIPMEGDEELKLMADFLETFYHGDEFGEKGIIMSYADEFEARVITWNIDPY 180

Db 121 LIDKGJLIPMEGDEELKLMADFLETFYHGDEFGEAEGPILMSYADEGARVITWNIDPY 180

Qy 181 VDVVSNSPREM1KRFVQ1VREKDPVDTIYTNGDNFDLPLYKRAEKLGVTLLGRDKEHP 240

Db 181 VDVVSNSPREM1KRFVQ1VREKDPVDTIYTNGDNFDLPLYKRAEKLGVTLLGRDKEHP 240

Qy 241 PKTHRMGDSFAVBIKGTHFDPYVTRTINLPTYLEAVYBAVIGKTKSKLGAEELAAI 300

Db 239 PKIQRMGDRFAVEVKGTHFDPYVTRTINLPTYLEAVYBAVQSPKRVAYEITPTA 298

Qy 301 WETEESMKKLAQYSMEDARATYELGKPEPMEAEELAKLIGOSWVDSRSSSTGNLYEWYL 360

Db 299 WETEESMKKLAQYSMEDARATYELGKPEPMEAEELAKLIGOSWVDSRSSSTGNLYEWYL 358

Qy 361 RAYERNEELAPNPKDECEYLARR-RQSTEGGYKPEPGLWNVNIVLFRSYPSTTIN 420

Db 359 RAYERNEELAPNPKDECEYLARR-RQSTEGGYKPEPGLWNVNIVLFRSYPSTTIN 417

Qy 421 VSDDTLEPEGKNDYAP1VGYKFCRDPFGIPSIGELITMRQETKRRKATIDIEKK 480

Db 418 VSDDTLEPEGKNDYAP1VGYKFCRDPFGIPSIGELITMRQETKRRKATIDIEKK 477

Qy 481 MLDYQRRAVKLHANSYTGYMGYPKARYSKCAGESTAWGHYIETIKEEKGFKVL 540

Db 478 LILYQRRAKILLANSYKQYGYTARARYCKEAESVTAWGHYIETIKEEKGFKVY 537

Qy 541 YADTDGFATATIPGKPKETIKKAKAKFVKYINGKLPGLELEYEGFYLRGFVAKKRYAVI 600

Db 538 YSDPDGFATIPGKPKETIKKAKAKFVKYINGKLPGLELEYEGFYLRGFVAKKRYAVI 597

Qy 601 DERTGTTTGCLEVRDQSRERAKTQVYKLEALKEDSVKAVETVQDVYELAKYQVPL 660

Db 598 DEEKSITTRGLELVRDQSRERAKTQVYKLEALKEDSVKAVIVTEKLSKVBPV 657

Qy 661 EKVLYTHEQITKOLSEYKAIGPHYIAIAKRLAAGKIKVRPGTISIYIVLGRSGSKISDRVILL 720

Db 658 EKVLYTHEQITKOLSEYKAIGPHYIAIAKRLAAGKIKVRPGTISIYIVLGRSGSKISDRVILL 717

Qy 721 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 773

Db 718 DEFPPTKHYDAEYYIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 774

Qy 731 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 775

Db 722 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 776

Qy 732 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 777

Db 723 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 778

Qy 733 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 779

Db 724 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 780

Qy 734 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 781

Db 725 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 782

Qy 735 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 783

Db 726 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 784

Qy 736 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 785

Db 727 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 786

Qy 737 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 787

Db 728 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 788

Qy 738 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 789

Db 729 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 790

Qy 739 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 791

Db 730 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 792

Qy 740 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 793

Db 731 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 794

Qy 741 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 795

Db 732 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 796

Qy 742 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 797

Db 733 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 798

Qy 743 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 799

Db 734 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 800

Qy 744 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 801

Db 735 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 802

Qy 745 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 803

Db 736 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 804

Qy 746 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 805

Db 737 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 806

Qy 747 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 807

Db 738 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 808

Qy 748 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 809

Db 739 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 810

Qy 749 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 811

Db 740 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 812

Qy 741 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 813

Db 742 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 814

Qy 743 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 815

Db 744 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 816

Qy 745 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 817

Db 746 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 818

Qy 747 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 819

Db 748 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 820

Qy 749 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 821

Db 750 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 822

Qy 751 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 823

Db 752 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 824

Qy 753 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 825

Db 754 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 826

Qy 755 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 827

Db 756 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 828

Qy 757 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 829

Db 758 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 830

Qy 759 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 831

Db 760 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 832

Qy 761 SEYDPKHKYDPPDVIENQVLPAVLRLAEGYRKEDILKQSSKQYGLDAWIK 833

Db 762 SEYDPKHKYDPPDVIENQVLPAVL

418 VSPDTLNREGCKEKEYDVAPOVGHRFCKDFPGTIPSLLGDLLEERQKIKKMKATIDPIERK 477
 Db 481 MLDYRORAKVLHANSYYGMPKARWYSKCEAESYNTAAGRHYLEMTEIEBKFGFKYL 540
 Qy 478 LDYRORAKVLHANSYYGMPKARWYCEAESYNTAAGRHYLEMTEIEBKFGFKYL 537
 Db 541 YADTDGFYATIPGKETIKEKAKEFKEYINSKLPGLLELEYEFPYLRGPFVAKRKYAVI 600
 Qy 538 YADTDGFATIPGKADATVCKRAMEFNTYNAKLPGALELEYEFPYLRGPFVAKRKYAVI 597
 Db 601 DEEGRITTRGLEYRDWSEETAKETOAKVLFALLKEDSVEKAVEIYKDVYEEIAKYQVSL 660
 Qy 598 DEEGRITTRGLEYRDWSEETAKETOAKVLFALLKEDSVEKAVEIYKDVYEEIAKYQVSL 657
 Qy 661 EKLVHQBQITDLSYEKAIGPHVALKRLANGKIKYRPGTISYVLRGGKISDRVIL 720
 Db 658 EKLVHQBQITDLSYEKAIGPHVALKRLANGKIKYRPGTISYVLRGGKISDRVIL 717
 Qy 721 SEMDPKHKYDDYIENQVLPAVRLLEAFAYKREDLQYKOSKQVGLDAWLK 773
 Db 718 DEEDPTKHKYDAFYIENQVLPAVRLLEAFAYKREDLQYKOSKQVGLDAWLK 770

RESULT 3
 US 08-656-005A-1
 Sequence 1, Application US/08656005A
 / Patent No. 6054301
 / GENERAL INFORMATION:
 / APPLICANT: KITABAYASHI, Masao
 / APPLICANT: ARAKAWA, Taku
 / APPLICANT: INOUE, Hiroaki
 / APPLICANT: KAWAKAMI, Bunsei
 / APPLICANT: KAWANTURA, Yoshihisa
 / APPLICANT: IMANAKA, Tadayuki
 / APPLICANT: TAKAGI, Masahiro
 / APPLICANT: MORTIWA, Masakazu
 / TITLE OF INVENTION: A Method of Amplifying Nucleic
 / NUMBER OF SEQUENCES: 16
 / NUMBER OF SEQUENCES: 16
 / ADDRESSSEE: Keyon & Kenyon
 / STREET: 1025 Connecticut Avenue, N.W., Suite 600
 / CITY: Washington
 / STATE: D.C.
 / COUNTRY: USA
 / ZIP: 20036
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: 3.5" Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: WordPerfect 5.1
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/656,005A
 / FILING DATE: 24 MAY 1996
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: JP 134096/95
 / FILING DATE: 31 MAY 1995
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Toffenetti, Judith L.
 / REGISTRATION NUMBER: 39,048
 / REFERENCE/DOCKET NUMBER: 2418/3
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 202-429-1776
 / TELEFAX: 202-429-0796
 / INFORMATION FOR SEQ ID NO: 1:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 774 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: double
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / US-08-656-005A-1

Query Match 80.0%; Score 3222.5; DB 3; Length 774;
 Best Local Similarity 77.4%; Pred. No. 2e-242; Index 3; Gaps 2;
 Matches 598; Conservative 83; Mismatches 89; Indels 3;

Qy 1 MTDITDTKDGKPIRIFKCKENGEFKEYDLPHPFQPYIYALLKDDSAIEKAIGSERHG 60
 Db 1 MLDITDITEDGPVIRIFKCKENGEFKEYDLPHPFQPYIYALLKDDSAIEKAIGSERHG 60
 Qy 61 KIVRVDVAKVCKKFLGRDVEWKL1FEHPOVPAVLRGKIREHPAVDIDYDIPPAKRY 120
 Db 61 TVTVTRVKEVKQKFLGRPVEWKL1FTHPOVPAVKDKEHGAVIDYDIPPAKRY 120
 Qy 121 LIDKG1PMEGDEEFLKLMADFETEYHEDFEGKGETIMISYADEBEARVITWKNDLDPY 180
 Db 121 LIDKG1VNEGDEEFLKLMADFETEYHEDFEGKGETIMISYADEBEARVITWKNDLDPY 180
 Qy 181 VDVYNSNEREMIKRPFQVIVREKOPDVLITYNGDNFDLDPYLIKRAEKUGVTLIGRDEKEHPE 240
 Db 181 VDVYSTEREMIKRFLRYVKEKDOPDVLITYNGDNFDAYLKERCEKIGINFALGRDGS-E 238
 Qy 241 PKIHMGSFAYB1KGR1HFDFPVTBTINP7YLEAVENAVLGKTKS1GAAE1AAI 300
 Db 239 PKIQRMGDRFATEVKGR1HFDLYPVIRTINTNP7YLEAVENAVLGKTKS1GAAE1AAI 298
 Qy 301 WETTB1SMRKLQYSMEDRATYELGRPFPPMELAKL1GQSVWDVRSSTGNLVENWYL 360
 Db 299 WEIGENLBRVARYSMEADAKVYELGRFELPNEAQSLWVNFNWL 358
 Qy 361 RVAYERNELAPNKPDEEFYRRLRTYLGGYTKEPGLWENITYLDFRCLYPSITVTHN 420
 Db 359 RKAYERNELAPNKPDEKLARR1PSTVEGGTYKEPGLWENIVYDFRSLYPSITVTHN 417
 Qy 421 VSPDTLREGCKNQVDAPIVGKFCDFPGFIPSISLGELITMROEIKKMKATIDPIEKK 480
 Db 418 VSPDTLREGCKEDVAQVGHFFCDFPGFIPSISLGELITMROEIKKMKATIDPIEKK 477
 Qy 481 MLDRYQRAVKLHANSYYGMPKARWYSKCEAESYNTAAGRHYLEMTEIYBKFQKYL 540
 Db 478 LLDRYQRATKILANSYYGMPKARWYSKCEAESYNTAAGRHYLEMTEIYBKFQKYL 537
 Qy 541 YADTDGFYATIPGKETIKKAKFELKYINSKLPGLLELEYEFPYLRGPFVAKRKYAVI 600
 Db 538 YSDTDGFATIPGDAEYTVKKKAMEFLNYINAKLPGLLELEYEFPYLRGPFVAKRKYAVI 597
 Qy 601 DEEGRITTRGLEYVRDSEIAKETOAKVLEALLKEDSVEKAVTVDVVEIATKQVPL 660
 Db 598 DEEGRITTRGLEYVRDSEIAKETOAKVLEALLKEDSVEKAVTVDVVEIATKQVPL 657
 Qy 661 EKLYTHEQITDOLSEYKAIGPHVALAAGKIKYRPGTISYVLRGSKGKISDRVILL 720
 Db 658 EKLYTHEQITDOLKDYKATGPHYAVARLAARGVKIRPGTVISYVLRGSKGIGDAAIPF 717
 Qy 721 SEYDPKHKYDDYIENQVLPAVRLLEAFAYKREDLQYKOSKQVGLDAWLK 773
 Db 718 DEFDPKHKYDAFYIENQVLPAVRLLEAFAYKREDLQYKOSKQVGLDAWLK 770

RESULT 4
 US-09-07-259-1
 / Sequence 1, Application US/09073259
 / Patent No. 614336
 / GENERAL INFORMATION:
 / APPLICANT: IMANAKA, Tadayuki
 / APPLICANT: TAKAGI, Masahiro
 / APPLICANT: MORIKAWA, Masaki
 / TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase
 / NUMBER OF SEQUENCES: 16
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Kenyon & Kenyon
 / STREET: 1025 Connecticut Avenue, N.W., Suite 600
 / CITY: Washington
 / STATE: D.C.

COUNTRY: USA
 ZIP: 20036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WordPerfect 6.1 Windows
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US-09/073,259
 FILING DATE: Concurrent Herewith
 CLASSIFICATION:
 APPLICATION NUMBER: 08/656,005
 FILING DATE: 24 MAY 1996
 APPLICATION NUMBER: JP 134096/95
 FILING DATE: 31 MAY 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Toffenetti, Judith L.
 REGISTRATION NUMBER: 39,048
 REFEERENCE/DOCKET NUMBER: 2418/10
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-429-1776
 TELEFAX: 202-429-0796
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 774 amino acids
 TYPE: amino acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-073-259-1

Query Match Score 3222.5; DB 3; Length 774;
 Best Local Similarity 77.4%; Pred. No. 2e-242;
 Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

Qy 1 MIFDTXYITRDGKPIIRIFKENGFEKFIELDPHFPOPIYALLKODSAIDBIKA-KGERHG 60
 1 MILDITYITBDGKPIIRIFKENGFEKFIELDPHFPOPIYALLKODSAIEVKITAEHG 60

Qy 61 KIVRVDVAVKVKRKEKFLGDKVEWKLLIPEHPODVAFLRGKIREHPIADYDIPFARRY 120
 61 TVVTRVVERVKQVKFLGDKVEWKLLIPEHPODVAFLRGKIREHPIADYDIPFARRY 120

Db 121 LIDKGLIPMEGDEFLKLMADFETYEGDBFGKGETIMISYADEEEARVITWNIDIPY 180
 121 LIDKGLIPMEGDEFLKLMADFIDCITYEGEBAEGPMLIYADEBEGARYITWNIDIPY 180

Qy 181 VDVSNSEREMIKRPFVQIYBEPKDPVLIITYGNDNFLLPYLIKRAEKGTVLLGRDKENHPE 240
 181 VDVSTEREMIKRPFVQIYBEPKDPVLIITYGNDNFLLPYLIKRAEKGTVLLGRDKENHPE 240

Db 241 PKIHRMGDSFAVEIKRPHIDLEPVTARTINPFTYIPEVAVENLICKTSKLGAEIAAI 300
 239 PKIHRMGDSFAVEIKRPHIDLEPVTARTINPFTYIPEVAVENLICKTSKLGAEIAAI 300

Qy 301 WETBBSMKLQYSMEDARATYLGKEFPMMAELAKLIGGSWDSRSSTGNLVEWYLL 360
 299 WETGENLERYARVYSMEDARATYLGKEFPMMAELAKLIGGSWDSRSSTGNLVEWYLL 358

Db 361 RVAYERNELAPNKEDEEYRRRLTITYGKEPEGLWENITYLDFRLCPLSYLITVTHN 420
 359 RKAYERNELAPNKEDEKELARR-RQSTEGGYKEPERGLWENIVYDFRSSTGNLVEWYLL 417

Qy 421 VSPDILREBCKNYDAPFVGZKFCDFGKPFPSILGLITRQEIKKMKATIDPIKK 480
 418 VSPDILREBCKNYDAPFVGZKFCDFGKPFPSILGLITRQEIKKMKATIDPIKK 477

Db 481 MLDRFRQAVKLHNSYTGNGYKARWYKSCAEASTAWGRAYIENIKE1BEKFGFVKL 540
 478 LLDRFRQRAKILANSYTGNGYKARWYKSCAEASTAWGRAYIENIKE1BEKFGFVKI 537

Qy 538 YSDTQFFATPGADAEYTKKAMFINTAKLPGABLEEGFYKRGFFVTKKAVI 597
 601 DEEGGTTTRGLEYIVRDSEIAKETQAKWLEAKEDSVEKAEIVYDVERIAKYOVPL 660
 598 DEEGGTTTRGLEYIVRDSEIAKETQARVLEAKDGVKAIRVTEKLSKEVPP 657
 661 EKLVTHEQTQTKDSEYKAGPHVAKLAAKGKIKWPTGTTISYVILRGSKGKISDRVILL 720
 658 EKLVTHEQTTRDLKQYKATGPHVAKLAARVKIRGTVTSYVILRGSKGKGRGDAIPF 717

Qy 721 SEYDPKHHKCYDPDYYIENQVLPAVRLAEGYKREDLKYQSSKQVGLDANK 773
 718 DEFDPTKHKYDAEYIENQVLPAVRLAEGYKREDLRYQTRQVGLISAWLK 770

Db 718 DEFDPTKHKYDAEYIENQVLPAVRLAEGYKREDLRYQTRQVGLISAWLK 770

RESULT 5
 US-09-363-095-1
 / Sequence 1, Application US/09363095
 / Patent No. 6187573
 / GENERAL INFORMATION:
 / APPLICANT: IMANAKA, Tadayuki
 / APPLICANT: TAKAGI, Masahiko
 / APPLICANT: MORIKAWA, Masakai
 / TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase
 / FILE REFERENCE: 2418/11
 / CURRENT APPLICATION NUMBER: US/09/363,095
 / CURRENT FILING DATE: 1999-07-30
 / EARLIER APPLICATION NUMBER: US 09/073,259
 / EARLIER FILING DATE: 1998-05-06
 / EARLIER APPLICATION NUMBER: US 08/656,005
 / EARLIER FILING DATE: 1996-05-24
 / EARLIER APPLICATION NUMBER: JP 134096/95
 / EARLIER FILING DATE: 1995-05-31
 / NUMBER OF SEQ ID NOS: 16
 / SEQ ID NO: 1
 / SOFTWARE: WordPerfect 6.1 Windows
 / LENGTH: 774
 / TYPE: PRT
 / ORGANISM: Hyperthermophilic archaeon
 US-09-363-095-1

Query Match Score 3222.5; DB 3; Length 774;
 Best Local Similarity 77.4%; Pred. No. 2e-242;
 Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

Qy 1 MIFDTXYITRDGKPIIRIFKENGFEKFIELDPHFPODVAFLRGKIREHPIADYDIPFARRY 60
 1 MILDITYITBDGKPIIRIFKENGFEKFIELDPHFPODVAFLRGKIREHPIADYDIPFARRY 60

Db 61 KIVRVDVAVKVKRKEKFLGDKVEWKLLIPEHPODVAFLRGKIREHPIADYDIPFARRY 120
 61 TVVTRVVERVKQVKFLGDKVEWKLLIPEHPODVAFLRGKIREHPIADYDIPFARRY 120

Qy 121 LIDKGLIPMEGDEFLKLMADFETYEGDBFGKGETIMISYADEEEARVITWNIDIPY 180
 121 LIDKGLIPMEGDEFLKLMADFIDCITYEGEBAEGPMLIYADEBEGARYITWNIDIPY 180

Db 181 VDVSNSEREMIKRPFVQIYBEPKDPVLIITYGNDNFLLPYLIKRAEKGTVLLGRDKENHPE 240
 181 VDVSTEREMIKRPFVQIYBEPKDPVLIITYGNDNFLLPYLIKRAEKGTVLLGRDKENHPE 240

Qy 241 PKIHRMGDSFAVEIKRPHIDLEPVTARTINPFTYIPEVAVENLICKTSKLGAEIAAI 300
 239 PKIHRMGDSFAVEIKRPHIDLEPVTARTINPFTYIPEVAVENLICKTSKLGAEIAAI 300

Db 301 WETBBSMKLQYSMEDARATYLGKEFPMMAELAKLIGGSWDSRSSTGNLVEWYLL 360
 299 WETGENLERYARVYSMEDARATYLGKEFPMMAELAKLIGGSWDSRSSTGNLVEWYLL 358

Qy 361 RVAYERNELAPNKEDEEYRRRLTITYGKEPEGLWENITYLDFRLCPLSYLITVTHN 420
 359 RKAYERNELAPNKEDEKELARR-RQSTEGGYKEPERGLWENIVYDFRSSTGNLVEWYLL 417

Db 481 MLDRFRQAVKLHNSYTGNGYKARWYKSCAEASTAWGRAYIENIKE1BEKFGFVKL 540
 478 LLDRFRQRAKILANSYTGNGYKARWYKSCAEASTAWGRAYIENIKE1BEKFGFVKI 537

Qy 538 YADTQFFATPGADAEYTKKAMFINTAKLPGABLEEGFYKRGFFVTKKAVI 597
 601 YADTQFFATPGADAEYTKKAMFINTAKLPGABLEEGFYKRGFFVTKKAVI 597

Qy	541	YADTQFYATIPGEKPEKKKAKAFKLTKINSCKGLELEYEGGLYLRGPVAKRYAVI	600
Db	538	YADTUGLHTAPGAETVKKKAKAFKLTKINSCKGLELEYEGGLYLRGPVAKRYAVI	597
Qy	601	DEEGRITTRGEVWRDWSEIAKEETQAKVLEAILEKDSVKEAVETKDVVEIAKYQVPL	660
Db	598	DEEGRITTRGEVWRDWSEIAKEETQAKVLEAILEKDSVKEAVETKDVVEIAKYQVPL	657
Qy	661	EKLVHEQITKDLSYKAIGPHVATAKRLAAGKIKVPGTISYTVLRGSGKISDRVILL	720
Db	658	EKLVHEQITRELKQYKATGPVATAKRLAARGKIRPGTIVSIVLKGSGRIGRAIPF	717
Qy	721	SEYDPRKHKYQDPTDYYTENQVLPAVLRLAFAFGYKEKDIKYQSSKQGIDAWL	772
Db	718	SEYDPRKHKYQDPTDYYTENQVLPAVLRLAFAFGYKEKDIKYQSSKQGIDAWL	769
RESULT 8			
	US-07-966-278-1		
	Sequence 1, Application US/07966278		
	Patent No. 5489523		
	GENERAL INFORMATION:		
	APPLICANT: Mathur, Eric A		
	TITLE OF INVENTION: EXONUCLEASE-DEFICIENT THERMOSTABLE		
	NUMBER OF SEQUENCES: 4		
	PROCOCCEUS FURIOSUS DNA POLYMERASE I		
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE: Thomas Fitting		
	STREET: 12226 High Bluff Road, Suite 300		
	CITY: San Diego		
	STATE: CA		
	ZIP: 92130		
	COMPUTER READABLE FORM:		
	COMPUTER: IBM PC Compatible		
	OPERATING SYSTEM: PC-DOS/MS-DOS		
	SOFTWARE: Patentin Release #1.0, Version #1.25		
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER: US/07/966,278		
	FILING DATE: 1993-12-26		
	CLASSIFICATION: 435		
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:		
	FILING DATE:		
	ATTORNEY/AGENT INFORMATION:		
	NAME: Fitting, Thomas		
	REGISTRATION NUMBER: 34,163		
	REFERENCE/DOCKET NUMBER: STG0113P		
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE: 619-792-3380		
	TELEFAX: 619-792-8477		
	INFORMATION FOR SEQ ID NO: 1:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH: 775 amino acid		
	TYPE: AMINO ACID		
	TOPOLOGY: linear		
	MOLECULE TYPE: protein		
	HYPOTHETICAL: NO		
	ANTI-SENSE: NO		
	US-07-966-278-1		
Query	Query Match	78.3%	Score 3151; DB 1; Length 775;
	Best Local Similarity	74.6%	Pred. No. 7.4e-237;
	Matches	576	Conservative 100; Nismatches 94; Indels 2; Gaps 1;
Qy	1 MIFDTDYITKDGKPIIRIFPKKENGGFKIELDPHFOPTYIAALLKDDSAIDETIKAIMGERHG	60	
Db	1 MILDGYITEGKPKTRFLPKKENGGFKIEUDRTFPYITALLRDSKIEVKKCTGERHG	60	
Qy	61 KIVRYVDAVKVKKPLGLRDYBVKWKLIFEHQDVPALRGKIREHPAVIDYEDPFAKRY	120	

APPLICATION NUMBER: US 07/803,627
 FILING DATE: 02-DEC-1991
 APPLICATION NUMBER: US 07/620,568
 FILING DATE: 03-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/657,073
 FILING DATE: 19-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/776,552
 FILING DATE: 15-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Fitting, Thomas
 REGISTRATION NUMBER: 34,163
 REFERENCE/DOCKET NUMBER: STG0100P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-792-3880
 TELEX/FAX: 619-792-8477
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 775 amino acids
 TYPE: amino acid
 TOPOLOGY: Linear
 MOLECULE TYPE: Protein
 HYPOTHETICAL: NO
 ANTI SENSE: NO
 US-08-424-921-1

Query Match 78.3%; Score 3151; DB 1; Length 775;
 Best Local Similarity 74.6%; Pred. No. 7,48-237;
 Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

Qy 1 MIFDDYITKDGPILIRIFKKEFKEFLDPHQQPYIALLKDSADEIRAIKGGERHG 60
 Db 1 MILDYDYYIEEGKVKIRLYKENGKPKLBHDRFRPYIALLRDSKEEVKKITGERHG 60
 Qy 61 KIVRVDAYRKVKKFLGRDVEVKKLIFEPHQDQVPLRKGKIREHPAVDIFYDIPFARY 120
 Db 61 KIVRIVDVEVKVKKFLGRPITVKKLHQPQDQPTIREKVRPAAVIFEYDIPFARY 120
 Qy 121 LIDKGKIPMEGDEEKLMAFDIETYHEGDEFKGKEIIMISYADEEARVITKNDLPHY 180
 Db 121 LIDKGKIPMEGDEEKLMAFDIETYHEGDEFKGKEIIMISYADEEARVITKNDLPHY 180
 Qy 181 VDVNSNEREMIKRFTVQIVREKDPMVITYGNGNFDLPLKRAEFLGTYLGRDKEHPE 240
 Db 181 VEVVSEREMIKRURIRIREKDPMIVTYGNGDSDFPLAKRAGLKLTRGDS-E 238
 Qy 241 PKIHRMGDSFAVEIKGRTHEDLPLPVWRTINLPTYLEAVYAVLGKTKSKEGAETAAI 300
 Db 239 PKMORIGDNTAVETKGRTHDLVHITITNLPTYLEAVYAVLGKTKSKEGAETAAI 298

Qy 301 WETBESMKKLAQYSMEDARATYELGKSFPPMELAKLIGQSYWDVRSSTSNLVIEWTLL 360
 Db 299 WESGENLVRVAKYSMEDAKATYELGKSFPPMELAKLIGQSYWDVRSSTSNLVIEWTLL 358

Qy 361 RVAVERNEIAPNPKDPEEYTRRRLRTYLGYYREPERGLWNTTYLDRCLYPSIIVTHN 420
 Db 359 RKAVERNEIAPNPKDPEEYTRRRLRTYLGYYREPERGLWNTTYLDRCLYPSIIVTHN 418

Qy 421 VSPDPLERGKQNYDVAPIVGKFCDFPGFIPSISLGELITRQEIKKOMKATIDPEBKK 480
 Db 419 VSPDPLERGKQNYDVAPIVGKFCDFPGFIPSISLGELITRQEIKKOMKATIDPEBKK 478

Qy 481 MLDYRQRAYKLHANSYYGMYGPKARWYKTYLGRDKEHPE 540
 Db 479 LLDYRQRAYKLHANSYYGMYGPKARWYKTYLGRDKEHPE 538

Qy 539 YIDDGGLATIPGSESEKKKALEFVYKINSKLPGLILEEGFYKRGFFYTKKRYAVI 598
 Db 539 DEEGRITRGLYVRRDSEIATEKTOAVKLEALIKEDEVERAKEEIVKDVVERIAKYOVPL 660
 Qy 598 DEECKVITGGLEIYTRDSEIATEKTOAVKLEALIKEDEVERAKEEIVKDVVERIAKYOVPL 658
 Db 661 EKLYTHEQITKDLSEYKAIGPHYAIAXKLAAGKIKYRGTIIISYIVLRSKSKISDRVILL 720
 Qy 659 EKLAIEQYTRPHEYKAIGPHYAVAKLAAGKIKYRGTIIISYIVLRSKSKISDRVILL 718
 Db 721 SEYDPKHKYKDPPYIYENQVLPAVURLEAFYKKELEKQSSKQVGLDML 772
 Qy 719 EYDPKHKYDAEYIENQVLPAVURLEAFYKKELEKQSSKQVGLDML 770

RESULT 9
 Sequence 9
 US-08-424-921-1
 Patent No. 5545552
 GENERAL INFORMATION:
 APPLICANT: Mathur, Purified Thermostable Pyrococcus
 TITLE OF INVENTION: FuriouS DNA Polymerase I
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Brigham & Fitting
 STREET: 12526 High Bluff Road, Suite 300
 CITY: San Diego
 STATE: CA
 COUNTRY: USA
 ZIP: 92130
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US 08/424,921
 FILING DATE: 19-APR-1995
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:

Qy 541 YADTDGFATIPGSEKPKENIKKAKAEFLKTYINSKLPGULEYEGFLRGFFYAKKRYAVI 600
 Db 539 YIDDDGLATIPGSESEKKKALEFVYKTYLGRDKEHPE 598
 Qy 601 DEEGRITRGLYVRRDSEIATEKTOAVKLEALIKEKQSSKQVGLDML 660

RESULT 10

US-08-556-355A-1

Sequence 1, Application US/08556355A

Patent No. 5866395

GENERAL INFORMATION:

APPLICANT: MATHUR, Eric J.

TITLE OF INVENTION: Thermostable Pyrococcus

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/556,355A

FILING DATE: 13-NOV-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/424,921

FILING DATE: 19-APR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/803,627

FILING DATE: 02-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/779,846

FILING DATE: 21-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/657,073

FILING DATE: 19-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/620,568

FILING DATE: 03-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: Barker, M. Paul

REGISTRATION NUMBER: 32,013

REFERENCE/DOCKET NUMBER: 04121.0004-02

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 775 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-556-355A-1

Query Match 78.3% Score 3151; DB 2; Length 775;

Best Local Similarity 74.6% Pred. No. 7.4e-23;

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

1 MIFDDYITKDGKPLTRIFKRNENGKFIELDPHFQPYIYALLKDDSAIDEIAIKIGERHG 60

Db 598 DEEGKVITRGLIEVERDWSIAKETQARVDETLXHGVDVEAVRIVKQKANYETLPP 658

Db 61 EKLVHEQITKDLSEBYKAIGPHVATAKRLAAGKIKVPGTIIISIVLFGSGKLSDRVILL 720

Qy 61 EKLAYEQTIRPLHVKAIQPHVATAKLAAGKVKPVGIVTIVLDRGPsSNRALA 718

Db 721 SEYDPKHKYKDPYXYYENQVLPAVRLAEGYRKEDLKYQSSKQVGLDAWL 772

Qy 721 SEYDPKHKYKDPYXYYENQVLPAVRLAEGYRKEDLKYQSSKQVGLDAWL 770

Db 719 EYDPKHKYKDAEYYENQVLPAVRLAEGYRKEDLRYQKTRQVGLTSWL 770

Db 1 MILDYYITEFGKPVRLFKENGKPKIEDRTPYIYALLRDSKIEVKKCTTGERHG 60

Qy 61 KIVRIVDVAVKYKKEPGRDVEVKLIFEPHQDPVLRGKTREHPAVIDYBEYDIFPAKY 120

Db 61 KIVRIVDVEVKYKKEPGRDVEVKLIFEPHQDPVLRGKTREHQDVTPLFPEYDIFPARRY 120

Db 121 LIDKGJLIPMGDEELKLMADIEETYHEGDEFGKBEIMISYADEBEARVITKWDIDLPY 180

Qy 121 LIDKGJLIPMGDEELKLMADIEETYHEGDEFGKBEIMISYADEBEARVITKWDIDLPY 180

Db 121 LIDKGJLIPMGSEELKILADIEETYHEGDEFGKBEIMISYADEBEARVITKWDIDLPY 180

Qy 181 VDVSNSEREMKRTQVIREKDPDYLITGDNFLPLILKRAFELGTVLGRKEKHE 240

Db 181 VEVVSSEREMKRFPLTIREKDPDLYTNGDSFEPFLAKRAEKGKLTIGRGS-E 238

Qy 241 PKIHRNGDSFAVEIKGRIFHDLPFVVRTRINLPPTLEAVEAVGKTKSKLGAEBIAAI 300

Db 239 PKMPTGDMTAVEVGRIFHLYKTRINLPPTLEAVEAVGKPKBKVYABIAKA 298

Qy 301 WETEESNKKLACYSMEDARATEYELGKEFFPMBAELAKLIGOSWVRSSTGNLVBWYLL 360

Db 299 WESGENLERVAKYSMEDAKATYELGKEFFPMETIQSLRVLGQPLWVDSRSSGNLVBWYLL 358

Qy 361 RYAYERNELAPKPDDEBEYRRLRTYLGYYVKEPERGLMENITYLDFRLCLYPSLITIN 420

Db 359 RYAYERNELAPKPDDEBEYRRLRTYLGYYVKEPERGLMENIVVLDRLALYPSLITIN 418

Qy 421 VSPDTLREGCKQNYDVAPIVGKFCFKDFG3PTIPS1GELTMROEKKKOMKATIDPIEK 480

Db 419 VSPDTLNLLEGCKNYDAPQVHKFKCD1GPTSLIGHLUEERQKLTKMKETODIEKI 478

Qy 481 MLDYRQBAVKLHANSYYGMYKPARWMSKCAESVTAWGRHYIEMTKEIJEKGFKVL 540

Db 479 LLDYRQAKIQLANSPFGYCYAKAWYCECAESVTAWGRKYIENWKEIJEKGFKVL 538

Qy 541 YADTDGFVATIGERPEPTIKKAKEFKLYINSKLPOLLELEYEGFLYLRGPVAKERAYI 600

Db 539 YIDTDGLYATIGERPEPTIKKAKEFKLYINSKLPOLLELEYEGFLYLRGPVAKERAYI 598

Qy 601 DEEGKVITRGLEVRRDSETAKETOAKVLEAILEDSEKAVEIYKDVYETIAKYQVPL 660

Db 599 DEEGKVITRGLEVRRDSETAKETOQARVLETILKHGDVEAVRLYKEVIOKLANYEIPP 658

Qy 661 EXLVTHEQ-ITKDLSEYKAIGPHVATAKRLAAGKIVYRPGTISYVLRGSKISKDEVILL 720

Db 659 EKLAIVEQTRPHYKAIGPHVATAKVLAKGKVKPGMIVGYVLRGDGPISNRAIL 718

Qy 721 SEYDPKHKYKDPYXYYENQVLPAVRLAEGYRKEDLKYQSSKQVGLDAWL 772

Db 719 EYDPKHKYKDAEYYENQVLPAVRLAEGYRKEDLRYQKTRQVGLTSWL 770

RESULT 11

US-07-803-627A-1

Sequence 1, Application US/07803627A

Patent No. 5948663

GENERAL INFORMATION:

APPLICANT: MATHUR, Eric J.

TITLE OF INVENTION: Purified Thermostable Pyrococcus

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEMORY TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.3.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/803.6277A

FILING DATE: 02-DEC-1991

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/776,552

FILING DATE: 14-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/657,073

FILING DATE: 19-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/620,568

FILING DATE: 03-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: Barker, M. Paul

REGISTRATION NUMBER: 32,013

REFERENCE/DOCKET NUMBER: 04121.0004-00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)408-4400

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 775 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-803-627A-1

Query Match

Best Local Similarity 78.3%; Score 3151; DB 2; Length 775;

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

Qy 1 MIFDTDYITKDGKPIIRPKKENGEFKIELDHFQPYIYALLKDDSAIDEIKAIGERHG 60

Db 1 MILDVDTTEEGKPVTRIFLKKENGKFKLEHDTRFRPYIYALRDDSKEVYKIGERHG 60

Qy 61 KIVRVVDAVKKKFLGDDVWKLLIPEYDIPFAKY 120

Db 61 KIVRIVDVKVKFLGKPIITWKLYLHQPVDIYDIPFAKY 120

Qy 121 LIDKGGLIMEGDEBKLNIAFDIETFYHEDFGKGEITIMISYADEEARVITWNIDLPY 180

Db 121 LIDKGGLIMEGEBEKLIAFDIETLYHGEETBGKGPIMISYADEEARVITWNIDLPY 180

Qy 181 VDVSNELEMIKRFVQITREKDPDVILTYNGDNFLFLYLIKRAEKGVTJLLGRKEHPE 240

Db 181 VEYSSSELEMIKRFVQITREKDPDVILTYNGDSDFPVLAKRAEKGKLTIGRGS--E 238

Qy 241 PKIHRMGMSFAVEIKGTHFLPFLPVYRTRINPPTYLEAVLGLKTSKLGABIAAI 300

Db 239 PNRQIGMTAEVKGRHFIDYHVTRINPPTYLEAVLGLKTSKLGABIAAI 298

Qy 301 WETBEESMKKLAQYSMEDARATELGKEFPMEABLKLIGSVDWSRSSTGNLVEWLL 360

Db 299 WESGENLTERVARYSMEDAKATELGKEFPMEIQLSRLVGFLWDVRSRSSTGNLVEWLL 358

Qy 361 RVAYERNEILAPNPKDPEBYYRRLRTYLGIVKPEPRGLWNTYLDFLCFLYPSIITVN 420

Db 359 RKEYERNEVAPNPKPSEBEYQRLRESYTGGVKEPEKGWENIVYLDFLALYPSIITVN 418

Qy 421 VSPDTLREBGGKNYDVAIIVGKFCFDPGFIPSILGELEITMREIKKOMKATDPIECK 480

Db 419 VSPDTLNEJEGCQNYDIAQVGHKFCFDPGFIPSILGELEITMREIKKOMKATDPIECK 478

Qy 481 MLDYRQAKVCHANSYGMCPKARWYSKCAEESTAANGRYHLEMTEKEKFGFKWL 540

Db 479 LLDYRQAKVAKLANSFVGGYAKAWYCKCAEESTAANGRYELVWKELEEKFGFKWL 538

Qy 541 YADTDGYATPGKEKETIKKAKAEPFLYNSKLEGFLLEFEYEGYLRGFFVAKKRYAVI 600

Db 539 YIDTDGLYATPGGSESEIKKAKAEPFLYNSKLEGFLLEFEYEGYLRGFFVAKKRYAVI 598

RESULT 12

US-09-244-889A-1

Sequence 1, Application US/09244889A

Patent No. 6489150

GENERAL INFORMATION:

APPLICANT: MATHUR, Eric J.

TITLE OF INVENTION: Purified Thermosable Pyrococcus

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

SEQUENCE DATA:

1 MIFDTDYITKDGKPIIRPKKENGEFKIELDHFQPYIYALLKDDSAIDEIKAIGERHG 60

1 MILDVDTTEEGKPVTRIFLKKENGKFKLEHDTRFRPYIYALRDDSKEVYKIGERHG 60

61 KIVRVVDAVKKKFLGDDVWKLLIPEYDIPFAKY 120

61 KIVRIVDVKVKFLGKPIITWKLYLHQPVDIYDIPFAKY 120

121 LIDKGGLIMEGDEBKLNIAFDIETFYHEDFGKGEITIMISYADEEARVITWNIDLPY 180

121 LIDKGGLIMEGEBEKLIAFDIETLYHGEETBGKGPIMISYADEEARVITWNIDLPY 180

181 VDVSNELEMIKRFVQITREKDPDVILTYNGDNFLFLYLIKRAEKGVTJLLGRKEHPE 240

VEYSSSELEMIKRFVQITREKDPDVILTYNGDSDFPVLAKRAEKGKLTIGRGS--E 238

241 PKIHRMGMSFAVEIKGTHFLPFLPVYRTRINPPTYLEAVLGLKTSKLGABIAAI 300

PNRQIGMTAEVKGRHFIDYHVTRINPPTYLEAVLGLKTSKLGABIAAI 298

WETBEESMKKLAQYSMEDARATELGKEFPMEABLKLIGSVDWSRSSTGNLVEWLL 360

WESGENLTERVARYSMEDAKATELGKEFPMEIQLSRLVGFLWDVRSRSSTGNLVEWLL 358

RVAYERNEILAPNPKDPEBYYRRLRTYLGIVKPEPRGLWNTYLDFLCFLYPSIITVN 420

RKEYERNEVAPNPKPSEBEYQRLRESYTGGVKEPEKGWENIVYLDFLALYPSIITVN 418

VSPDTLREBGGKNYDVAIIVGKFCFDPGFIPSILGELEITMREIKKOMKATDPIECK 480

VSPDTLNEJEGCQNYDIAQVGHKFCFDPGFIPSILGELEITMREIKKOMKATDPIECK 478

MLDYRQAKVCHANSYGMCPKARWYSKCAEESTAANGRYHLEMTEKEKFGFKWL 540

LLDYRQAKVAKLANSFVGGYAKAWYCKCAEESTAANGRYELVWKELEEKFGFKWL 538

YADTDGYATPGKEKETIKKAKAEPFLYNSKLEGFLLEFEYEGYLRGFFVAKKRYAVI 600

YIDTDGLYATPGGSESEIKKAKAEPFLYNSKLEGFLLEFEYEGYLRGFFVAKKRYAVI 598

Query Match

Score 3151; DB 4; Pred. No. 7.4e-237;

Best Local Similarity 78.3%; Score 3151; DB 4; Pred. No. 7.4e-237;

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

Query 1 MIFDIDYTKDGKPIRKFKENGEFKIELDPHPQYIYALLKODSAIDEKAIAKGPRHG 60
 Database 1 MILDYDYTEEGPKVIRPKKENSKFKTEHDTRPYYALRDDSKKEEVKTKTGRHG 60

Query 61 KIVRVDAAVKVKKPLGRDVEWKLILFEPQDVALRGKIREPAVILYEDIPFKARY 120
 Database 61 KIVRVDVERKVKPLGKPTVWKLYLHEPQDVTIREXREHPAVVILFEDIPFKARY 120

Query 121 LIDKGILPPMGEDEBKLMAPDIEETYHEDEFGKGEIIMSYADEEARVITWNIDLPY 180
 Database 121 LIDKGILPPMGEDEBKLMAPDIEETYHEDEFGKGEIIMSYADEEARVITWNIDLPY 180

Query 181 VDVVSNEREMITKRFYQIVREKDPDYLITYNGDNFDLPLVYIILKRLGTYLILLGRDKHPE 240
 Database 181 VEVSSSEREMITKFRITRKDPLIVTYNGDSDPFLPLAKRALKGKLTKTGRGS-E 238

Query 241 PKIHRMGDSFAVEIKRHFDFLPVTRVTRNLPTYLTAEVAKLGTKSUGAAEIAAI 300
 Database 239 PRMQRGDMDTAEVGRHFDPLHXTTRTINLPTYLTAEVAKLGTKSUGAAEIAAI 300

Query 301 WETEESMKKLQASMEDARATYELCKEFFPMEEAELAKLJIGQSYMDVSSSSTGNLNEVYLL 360
 Database 299 WESGENINTERVAKYSMEDAKATYELCKEFFPMEEIQLSLRQPLMDVSSSSTGNLNEVYLL 358

Query 361 RVAYENELAQNPKDDEYERRLRITYLGGYKEBERGLWENITYLDPCLYSITIWHN 420
 Database 359 RKAYEAEVANKPSSEYQRRLRSYTGFKVKEBEKGWENITYLDPALYPSITIWHN 418

Query 421 VSPDTLREGEKKNYDAPIVGKFCDFPFFPSITLGELEITMREIJKKMKATIDPIKK 480
 Database 419 VSPDTLNLEGCKNYDAPQVGHFKCDIDGPIPSLGHILEERQKIKTMKETQDPIKKI 478

Query 481 MLDYRDAVLHANSYYGNGYPKARWYSKCBASVTAANGRYHETMIGEIEKEFGFVTL 540
 Database 479 LLDYRQAKIULLANSFYYGYYGAKARWYCKECAESVTAANGRYKTELWKELEEKFGFVTL 538

Query 541 YADTDGTYATIPGKEKPTIEKAKETLYKINSKPLGLLEBEYEGYLRFEVAKRVAI 600
 Database 539 YIDTDGJYATPGGSEIICKKALETVKVKINSKPLGLLEBEYEGYKRFEVAKRVAI 598

Query 601 DEEGRITRGLVREWWSEAKETOAKVLEALKEDSVEKAEVKDVVEEIAKYQVPL 660
 Database 599 DEEGKVTRGLBIVRWWSEAKETOAKVLEALKEDSVEKAEVKDVVEEIAKYQVPL 658

Query 661 EKLVTHQITKDLSEYKTAIGPHVATIAKRLAAGKIKVPRGPTIISYIYLRSGSKKSDRVIL 720
 Database 659 EKLVTHQITRHLSEYKTAIGPHVATIAKRLAAGKIKVPRGPTIISYIYLRSGSKKSDRVIL 718

Query 721 SEYDPKHKYDDYIINQNLPAVILAEAGYRKEDLKQSSKQYGLDawl 772
 Database 719 EYDPKHKYDAEYYTINQNLPAVILAEAGYRKEDLKQTRGTLTSWL 770

RESULT 613
 Database US-08-688-649-37
 ; Sequence 37, Application US/08688649
 ; GENERAL INFORMATION:
 ; APPLICANT: MAMONE, JOSEPH A.
 ; TITLE OF INVENTION: MODIFIED POL-11 TYPE DNA
 ; NUMBER OF SEQUENCES: 38
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U. S. A.
 ; ZIP: 90071-2066
 ; COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
 MEDIUM TYPE: storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/688,649
 FILING DATE: Filed Herewith
 CLASSIFICATION: 435
 PRIORITY APPLICATION NUMBER:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Warburg, Richard J.
 REGISTRATION NUMBER: 32,327
 REFERENCE/DOCKET NUMBER: 220/281
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 INFORMATION FOR SEQ ID NO: 37:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 776 amino acids
 TYPE: amino acid
 STRANDBEDES: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-688-649-37
 Query Match 78.3%; Score 3151; DB 2; Length 776;
 Best Local Similarity 74.6%; Pred. No. 7.4e-237;
 Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;
 Query 1 MIFDIDYTKDGKPIRKFKENGEFKIELDPHPQYIYALLKODSAIDEKAIAKGPRHG 60
 Database 1 MILDYDYTEEGPKVIRPKKENSKFKTEHDTRPYYALRDDSKKEEVKTKTGRHG 60
 Query 2 1 MLDYRDAVLHANSYYGNGYPKARWYSKCBASVTAANGRYHETMIGEIEKEFGFVTL 540
 Database 1 MLDYRQAKIULLANSFYYGYYGAKARWYCKECAESVTAANGRYKTELWKELEEKFGFVTL 538
 Query 3 61 KIVRVDAAVKVKPLGKPTVWKLYLHEPQDVPTRKIREHPLGKIREHPLDIFPAKRY 120
 Database 61 KIVRVDVEKVKPLGKPTVWKLYLHEPQDVPTRKIREHPLGKIREHPLDIFPAKRY 120
 Query 4 61 KIVRVDAAVKVKPLGKPTVWKLYLHEPQDVPTRKIREHPLGKIREHPLDIFPAKRY 120
 Database 61 KIVRVDAAVKVKPLGKPTVWKLYLHEPQDVPTRKIREHPLGKIREHPLDIFPAKRY 120
 Query 5 121 LIDKGILPPMGEDEBKLMAPDIEETYHEDEFGKGEIIMSYADEEARVITWNIDLPY 180
 Database 121 LIDKGILPPMGEDEBKLMAPDIEETYHEDEFGKGEIIMSYADEEARVITWNIDLPY 180
 Query 6 121 LIDKGILPPMGEDEBKLMAPDIEETYHEDEFGKGEIIMSYADEEARVITWNIDLPY 180
 Database 61 KIVRVDAAVKVKPLGKPTVWKLYLHEPQDVPTRKIREHPLGKIREHPLDIFPAKRY 120
 Query 7 121 LIDKGILPPMGEDEBKLMAPDIEETYHEDEFGKGEIIMSYADEEARVITWNIDLPY 180
 Database 121 LIDKGILPPMGEDEBKLMAPDIEETYHEDEFGKGEIIMSYADEEARVITWNIDLPY 180
 Query 8 181 VDVVSNEREMITKRFQIVREKDPDYLITYNGDNFDLPLVYIILKRLGTYLILLGRDKHPE 240
 Database 81 VDVVSNEREMITKRFQIVREKDPDYLITYNGDNFDLPLVYIILKRLGTYLILLGRDKHPE 240
 Query 9 181 VEVSSSEREMITKFRITRKDPLIVTYNGDSDPFLPLAKRALKGKLTKTGRGS-E 238
 Database 91 VEVSSSEREMITKFRITRKDPLIVTYNGDSDPFLPLAKRALKGKLTKTGRGS-E 238
 Query 10 121 PKIHRMGDSFAVEIKRHFDFLPVTRVTRNLPTYLTAEVAKLGTKSUGAAEIAAI 300
 Database 121 PKIHRMGDSFAVEIKRHFDFLPVTRVTRNLPTYLTAEVAKLGTKSUGAAEIAAI 300
 Query 11 239 PRQRIGDMTAEVKPKRHFPLHVTTRNLPTYLTAEVAKLIGOSVWYDYSRSTGNYLVEWYLL 360
 Database 101 WETEESMKKLQASMEDARATYELCKEFFPMEEAELAKLIGOSVWYDYSRSTGNYLVEWYLL 360
 Query 12 299 WESGENLERVAKYSMEDAKATYELCKEFFPMEEIQLSLRVQPLMDVSSSTGNLNEWFL 358
 Database 11 361 RVAYERNELAQNPKDDEYERRLRITYLGGYKEBERGLWENITYLDPCLYSITIWHN 420
 Query 13 359 RAYEAEVANKPSSEYQRRLRSYTGFKVKEBEKGWENITYLDPALYPSITIWHN 418
 Database 14 419 VSPDTLNLEGCKNYDAPQVGHFKCDIDGPIPSLGHILEERQKIKTMKETQDPIKKI 478
 Query 15 479 LLDYRQAKIULLANSFYYGYYGAKARWYCKECAESVTAANGRYKTELWKELEEKFGFVTL 538
 Database 16 491 MLDYRDAVLHANSYYGNGYPKARWYSKCBASVTAANGRYHETMIGEIEKEFGFVTL 540
 Query 17 491 MLDYRQAKIULLANSFYYGYYGAKARWYCKECAESVTAANGRYKTELWKELEEKFGFVTL 540
 Database 18 479 LLDYRQAKIULLANSFYYGYYGAKARWYCKECAESVTAANGRYKTELWKELEEKFGFVTL 538
 Query 19 541 YADDGFTATIPSEKPTIPEKFKYDYLQVTRGKTRGTLTSWL 770
 Database 20 539 YIDTDGLYATIPSEKPTIPEKFKYDYLQVTRGKTRGTLTSWL 770

Qy	601	DEGRITRGLERYRVDSEIATEKTOAVKLEAIIKEDOSVEKAVEIVRDVEEIAKYQVPL	666
Dy	599	DEGVKTRGLERYRVDSEIATEKTOAVKLEAIIKEDOSVEKAVEIVRDVEEIAKYQVPL	658
Qy	661	EXLIVHEPITDLSSEYKAIGPHVIAKRLAAGKIVRPGTILISYIVLRSGKISDRVILL	720
Dy	659	EXLIAVEQITRLHEYKAGPHVAVAKLAAKGKIVRPGTILISYIVLRSGKISDRVILL	720
Qy	721	SEYDPKHKYDPPDYIENQVPLPAVLRLTEAEGYRKEDLKYQSSKQYGLDAML	772
Dy	719	SEYDPKHKYDAAEYIENQVPLPAVLRLTEAEGYRKEDLRYQXTRQYGLTSML	770
RESULT 14			
US-09-585-858-35			
Sequence 35, Application US/09585858			
Patent No. 6492161			
GENERAL INFORMATION:			
APPLICANT: Sigríður Hjörleifsdóttir			
Gudmundur O. Hrggvidsdóttir			
APPLICANT: Olafur H. Fridjónsson			
APPLICANT: Arniður Ásvarsdóttir			
APPLICANT: Jakob K. Ákrísson			
TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic			
TITLE OF INVENTION: Host Organism			
FILE REFERENCE: 2739_1001-001			
CURRENT APPLICATION NUMBER: US/09/585, 858			
CURRENT FILING DATE: 2000-12-18			
PRIOR APPLICATION NUMBER: 60/137, 120			
PRIOR FILING DATE: 1999-06-02			
NUMBER OF SEQ ID NOS: 73			
SOFTWARE: FastSEQ for Windows Version 4.0			
SEQ ID NO 35			
TYPE: PRT			
ORGANISM: Desulforoccocus strain Tok			
US-09-585-858-35			
Query Match 78.0%; Score 3141.5; DB 4; Length 773;			
Best Local Similarity 74.6%; Pred. No. 4e-236; Gaps 2;			
Matches 577; Conservative 97; Mismatches 96; Indels 3;			
Qy	1	MIFDIDITRQKPRIRIYRCENGEERIYDHPQVPAIRGKIREHPAVDIDEYDIFPARY	60
Dy	1	MILDADYITDEGKPVIRFEKEKGKEDYIDRDFEYIYALKDOSAIEJKKTAERHG	60
Qy	61	KIVRYYDAVTKYKKEKGLRDVEWYKLIYFHEHPDQVPAIRGKIREHPAVDIDEYDIFPARY	120
Dy	61	TTVRVTRAEVYKKEKGLRDVEWYKLIYFHEHPDQVPAIRGKIREHPAVDIDEYDIFPARY	120
Qy	121	LIDKGKLIPMEDSEELKIMAFDIEYTYHGDBFGKGEIIMISYADEEARYITWMDLPLY	180
Dy	121	LIDRGKLIPMEDSEELKIMAFDIEYTYHGDBFGKGEIIMISYADEEARYITWMDLPLY	180
Qy	181	VDVVGNEREMIKRFQIVREKDPDYLITTYDNLFLPILYKRAEKLGLGTVLJLGRKEHPE	240
Dy	181	VESVSTEKEMIKRFQIVREKDPDYLITTYDNLFLPILYKRAEKLGLGTVLJLGRKEHPE	238
Qy	241	PKIHRGQDSFAVEIKRGIHFDLFPYVYRRTINLPTTYLEAVIYKTKSKGAEIETAI	300
Dy	239	PKIQRGQDSFAVEIKRGIHFDLFPYVYRRTINLPTTYLEAVIYKTKSKGAEIETAI	300
Qy	301	WETEBSMKKLAQYSMEDARAYELGKEFFMEAEIAKLIGOSVNDYVRSRSTGNYLVEVYLL	360
Dy	299	WESGEGLERVARYSMSMEDARAYELGKEFFMEAQSLRVLQSQLWVRSRSTGNYLVEVYLL	358
Qy	361	RVAYERNELAIPNKPDEEEYRPLTYLGGYKKEBERGLWENTYLDRCFLYPSIILVHN	420
Dy	359	RKAYERNDVAIPNKPDEEEYRPLTYLGGYKKEBERGLWENTYLDRCFLYPSIILVHN	417
Qy	421	VSPDTLREGKKNYDVAIPYGYKFCDFPGTIPSIGELITMROTKKMKATIDPIEK	480
Dy	418	VSPDTLREGKKNYDVAIPYGYKFCDFPGTIPSIGELITMROTKKMKATIDPIEK	477

QY 481 MLDYRQRAVKLHANSYYGMPYKARWYSKCSEAEEVTANGRHYLEMTEKIEEEKFGFKV1 540
 QY :
 Db 478 LLDYRQRAIKLANSYYGTYAYANRWCYRECAEEVTANGRQYETTREIEEKFGFKV1 537
 QY 541 YADTDGFYATIPGEKPEKIKKAKFELKTYNKSNUPLGLBLEYEGFYLRGFPVAKRYAVI 600
 QY :
 Db 538 YADTDGFATIPGAEETVTKNAKEFLNTINPRLGLBLEYEGFYRQGFFVTKKYAVI 597
 QY 601 DEEGRTTRGLEEVDRDWSIAKETQAKYEALEALKEDSVEKAEIVDVEEATKXQYPL 660
 QY :
 Db 598 DEEDDTTRGLEEVDRDWSIAKETQARYLEALKHGDEEAIRVKVETEKLSRHEYPP 657
 QY 661 EKLVHEQITRDKLSSYRAIGPHVAKLRAAKGKIVKRPTTISIVLRSGK1SDRVIL 720
 QY :
 Db 658 EKLVHEQITRDLRSYRATCPHVAKLRGK1KRPVIVSLKGPGRVGDRA1PP 717
 QY 721 SBYDPKKHVDPDTYENQVLPAVLRLIAFGYRKEDLYQKQVGLDAWLK 73
 QY :
 Db 718 DEFDPAKRYDAETIENQVLPAVERILRAFGRKEDLYQKTKQAGLGAWLK 770

RESULT 15
 US-08-375-134-12
 / Sequence 12, Application US/08375134
 / Patent No. 5602011
 / GENERAL INFORMATION:
 / APPLICANT: Luhm, Robert A.
 / APPLICANT: d' Hennezel, Olga B.
 / APPLICANT: Duffaud, Guy D.
 / APPLICANT: Jolly, James F.
 / APPLICANT: Kelly, Robert M.
 / APPLICANT: Ting, Eve Y.-Fay
 / TITLE OF INVENTION: PURIFIED THERMOCOCCUS BAROSSII DNAPOLYMERASE
 / NUMBER OF SEQUENCES: 13
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Quarles & Brady
 / STREET: 411 East Wisconsin Avenue
 / CITY: Milwaukee
 / STATE: Wisconsin
 / COUNTRY: U.S.A.
 / ZIP: 53202-4497
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patentin Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/375.134
 / FILING DATE: 08/08/1995
 / CLASSIFICATION: 435
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Baker, Jean C.
 / REGISTRATION NUMBER: 35,433
 / REFERENCE/DOCKET NUMBER: 740211.90440
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (414) 277-5709
 / TELEFAX: (414) 271-3552
 / INFORMATION FOR SEQ ID NO: 12:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 779 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 US-08-375-134-12

Query Match 78.0%; Score 3139; DB 1; Length 779;
 Best Local Similarity 76.2%; Pred. No. 6.4e-26;
 Matches 589; Conservative 81; Mismatches 99; Indels 4; Gaps 3;
 QY 1 MIFDTDXITKDGKPIIRFKKENGEKFIELDPHQPYIYALKDDAIDEKAIKGRHG 60

Db 1 MILDVYDITEDGKPVIRVFKCKGFKLBYDREFEPYIYALLRDDSABEELEKITAERHG 60
 Qy 61 KIVRYVDAVKVKCFGLRDEVWKLIFERHQDFPALR-GKIREHPAVIDIYBYDIPFAKR 119
 Db 61 KVVKYKRAEKVKCKPLGRSVWVLYFTHPDVPAIRDKIRGPAVIDIYDIPFAKR 120
 Qy 120 YLDIGLIPMEGDBELKLAUDIEFYHIGDEFERGETIMISTADEBARVITWNKIDLP 179
 Db 121 YLDIGLIPMEGDBELKLAUSFDELYHIGEEFTGPILMISTADEBARVITWKCIDLP 180
 Qy 180 YDVYSSNEREMIKREVOQIVREKDPVLYTNGDNFDLPLIKRAEKLGVTLLGRDKEHP 239
 Db 181 YDVVSTEREKIYKLKVVERDOPVLIYDGFDFAYLKKCEKLGVSFTLGRDSS-- 238
 Qy 240 EPKIHRMGDSFAVEIKGRTHFDLFPVYRRTINLPYTLEAVYRAVLGKTKSKIGAEETAA 299
 Db 239 EPKIQMGRDFAVEYGRTHFDLPAIRRTINLPYTLEAVYEAVFGPKEKYYAEEAT 298
 Qy 300 IWETESMKLQLQSMDARATYELGKEFPMELAKLIGQSWDVSRSSTONLWVYL 359
 Db 299 AWETGEGLEGARYSMDARVYELGREFPMPMAQLSRJLGGWMDVSRSSTGNLWVFL 358
 Qy 360 LRVAYERNELAPNKBDDEYRRRTTYYGYYKEPERGLWNTYDLCRCLYPSIIVTH 419
 Db 359 LRVAYERNELAPNKBDDEYLLARR-RGGYAGGGYKEPERGLWNTYJFRSLYPSIITH 417
 Qy 420 NVSPDTLREGCKNYDVAPIVGKFCDFGFGFISLGELETWQEIICRMATIDPEK 479
 Db 418 NVSPDTLNRECKSDVAPOVGHRECKDFGFGFISLGNLLEERQKICRMATIDPEL 477
 Qy 480 RMLDYZRQAVLHANSYYGNGYKARWYSKEACESTAWGRHYTEMITKEIEBFGRKV 539
 Db 478 KLLDYZRQAKILANSFGYGYARAWTCKEAEVSITAWGREYTEMTRLESEKFGKD 537
 Qy 540 LYADTDFGYATTIGKEPKETIKKAKFALKYINSKLPGGLELEYEGFYLRGFFAKKYYAV 599
 Db 538 LYADTFLHATIGADREYKCKDLEFLNVTNPKLGLBLELEYEGFYSRGFFVTKKAV 597
 Qy 600 IDEEGRTTTRGLEYVRDWSIETAKTOAKYLEATLKEDSYEKAETIVKDVVEETAKYCP 659
 Db 598 IDEEGRTTTRGLEYVRDWSIETAKTLLARYTEAHLRGDVEEATRIVVETEKLSKYCP 657
 Qy 660 LEKLVTHEQITKDLSEYKAIGPHVIAKRLAAGJKVPRDTISYIYLRGSKKISDRVIL 719
 Db 658 PEKLVTEQITRELYKATGPHVIAKRAARGKIRPSTVIVKGSGRGDRAP 717
 Qy 720 LSEYDPKHKYDPTTENQVLPATLRAFGYRKEDLKYQSSKQVGDawl 772
 Db 718 EDEFDPKHRDADYIENQVLPATLRAFGYKKEDEYQKTRQVGLGawl 770

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - Protein search, using SW mode!

Run on: February 9, 2004, 11:46:55 ; Search time 40 Seconds
 (without alignments)

18660.864 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026

Sequence: 1 MIFPTDYTKDGKPIIRIFK.....KEDLKYQSSKQVGLDAWKK 774

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 98168662 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing First 45 Summaries

Database : PIR76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3356	83.4	1702	2 S42459	DNA-directed DNA polymerase Vent
2	3204.5	79.6	775	2 S67920	DNA-directed DNA polymerase I
3	3164.5	78.6	771	2 C75023	DNA-directed DNA polymerase I
4	3151.3	78.3	775	2 S35543	DNA-directed DNA polymerase I
5	2944.5	72.9	1312	2 S68593	DNA-directed DNA polymerase I
6	2886	71.7	1235	2 C71210	Probable DNA-directed DNA polymerase
7	2764.5	68.7	1670	2 S71551	DNA-directed DNA polymerase B1
8	1301	32.3	781	2 A69312	DNA-directed DNA polymerase B1
9	1291	32.1	1634	2 E64410	DNA-directed DNA polymerase B1
10	1209.5	30.0	586	2 C69028	DNA-dependent DNA polymerase
11	1204.5	29.9	781	2 JC7382	DNA-directed DNA polymerase
12	1190	29.6	803	2 B56277	DNA-directed DNA polymerase
13	1138	28.3	784	2 E72515	Probable DNA-directed DNA polymerase
14	693.5	17.2	1088	2 T05731	DNA-directed DNA polymerase
15	676.5	16.8	764	2 S75407	Probable DNA-directed DNA polymerase
16	659.5	16.4	1038	1 JC5757	DNA-directed DNA polymerase
17	657.5	16.3	1038	2 T18222	DNA-directed DNA polymerase
18	657.5	16.3	1086	2 T40242	DNA polymerase delta
19	649.5	16.1	1086	2 T43266	DNA-directed DNA polymerase B1
20	626	15.5	901	2 E84210	DNA-directed DNA polymerase B1
21	621.5	15.4	1097	1 RNBYL3	DNA-directed DNA polymerase B1
22	619.5	15.4	882	2 S23019	DNA-directed DNA polymerase B1
23	619.5	15.4	1084	1 S19661	DNA-directed DNA polymerase B1
24	616.5	15.3	872	2 JC7380	DNA polymerase I
25	616.5	15.3	882	2 F90243	DNA-directed DNA polymerase I
26	614.5	15.3	1105	1 S22573	DNA-directed DNA polymerase I
27	614.5	15.3	1094	2 A41618	DNA-directed DNA polymerase I
28	611	15.2	1107	1 T28153	Probable DNA-directed DNA polymerase I
29	609	15.1	1513	2 T28158	Probable DNA-directed DNA polymerase I

Query Match 83.4% ; Score 3355; DB 2; Length 1702;
 Best Local Similarity 42.7%; Pred. No. 2-164;
 Matches 33; Mismatches 726; Conservative 15; Indels 928; Gaps 2;
 Qy 1 MIFDTDITXDGKPIIRIFKENGKFRKIELDPHPQPYIALLKDSADIBIAKKGHRG 60

Db	1 MILDITYITKDGKPITRIFKKEENGKFKEILDPHQPYITALLKDSAIEBIAIKGEHH	60	Db	1081 DSVGESEIIIRONGKIRFKVVKIDLFLSKVDSIGBKEYCITLEGYEALTHLDDGKLYWKPV	1140	
Qy	61 KIVRYVDVAKVKKKFGLRDYEVWKLIFEHPODPALRGKIREHPDAVIDEYDIPFARY	120	Qy	544	-----	
Db	61 KIVRYVDVAKVKKKFGLRDYEVWKLIFEHPODPALRGKIREHPDAVIDEYDIPFARY	120	Db	1141 PYVMRHRANKDMERTWITNSMYWIDYTEDHSUIGINTSKTKTAKIGERLKEVKPFELGK	1200	
Qy	121 LIDKGLIPMEGDEELKLMAFDIETFYHGDEFGKGEIIMISYADEEAVITWTKNIDLPY	180	Qy	544	-----	
Db	121 LIDKGLIPMEGDEELKLMAFDIETFYHGDEFGKGEIIMISYADEEAVITWTKNIDLPY	180	Db	1201 AVKSLICPNAPLKDENTKTSEIAVKFWELVGLIVDGDNWGGDSRKAEYIYLGLSTGKDAAE	1260	
Qy	181 DVVSNEREMKRFQIVREXPDVYLITYGDNFLPILYIKRAEKLGLVTLIGRDKEPE	240	Qy	544	-----	
Db	181 DVVSNEREMKRFQIVREXPDVYLITYGDNFLPILYIKRAEKLGLVTLIGRDKEPE	240	Db	1261 IKOKLBLEPLKTYGVISNYYPKNEKGDENILAKSLIKFMRKHFDEKGKRIKPEMYELPV	1320	
Qy	241 PKIHRMGDSFAVEIKRIFHFLPILFVPRTRNLNPLTYLAEAVLGKTKSKLGAEELAAI	300	Qy	544	-----	
Db	241 PKIQRMGSFAVEIKRIFHFLPILFVPRTRNLNPLTYLAEAVLGKTKSKLGAEELAAI	300	Db	1321 TYIEAFDRLGLFSADGTVTIRKGVPETRLTNIDAFIREVRLKLWVIGNSNSIAPTTPNR	1380	
Qy	301 WETEESMKLQASMEDARATYELGKFFPMEAEELAKLICQSVWVRSSTGNLVEWTL	360	Qy	544	-----	
Db	301 WETEESMKLQASMEDARATYELGKFFPMEAEELAKLICQSVWVRSSTGNLVEWTL	360	Db	1381 YNGVSTGTYSKHLIRKNKWRFAERIGFLIERKQRILLEHKSARVTRNTIDGFFDLVHVK	1440	
Qy	361 RAYERELAPKPKDEEYERERLRTTYLGTYKEPRLGLVNTITYLFRCLYPSLIVTHN	420	Qy	544	-----	
Db	361 RAYERELAPKPKDEEYERERLRTTYLGTYKEPRLGLVNTITYLFRCLYPSLIVTHN	420	Db	1441 KVEEIPYEGYYVDIEVEETHFFFANNLVLTNTDGFATIPEKPKLKKAKEFELKYINS	572	
Qy	421 VSPDTLRECGCKNVDYAPIVGKFKCFPGTIPSILGELTMROBIKGMKATIDPIECK	480	Qy	573 KLPGLLLEYEGRYLRFFVAKGRYAVIDEGRITTRGLEYVRWOWSEIAKETQAKYLEA	632	
Db	421 VSPDTLRECGCKNVDYAPIVGKFKCFPGTIPSILGELTMROBIKGMKATIDPIECK	480	Db	1501 KLPGLLLEYEGRYLRFFVAKGRYAVIDEGRITTRGLEYVRWOWSEIAKETQAKYLEA	1560	
Qy	481 MLDYRQRAVHLA-----	493	Qy	633 ILKEDSYEKALEVQDVEELAKYQFPLEKUHIEQITQDSEYKAGIPHYIAKRLAAK	692	
Db	481 MLDYRQRAKLLANSILPNEWLPPIEENGIEIKFVKIGBFINSYMEOKEQKENVTKVENTEVL	540	Db	1561 ILKEGSYEKALEVQDVEELAKYQFPLEKUHIEQITQDSEYKAGIPHYIAKRLAAR	1620	
Qy	494 -----	493	Qy	693 GIKVRPGTISIYIVLRSGKISDRVILLSEYDPKHYDPPYIYENQVLPAVRLIBAFG	752	
Db	541 VNNLFAFSFNKKIKKESSFVKVKAIRHKYKGKAYEIQLOSSGRKINTAGHSLFTVRNGEI	600	Db	1621 GIKVPGIIISITVILKSGKISDRVILLSEYDPKHYDPPYIYENQVLPAVRLIBAFG	1680	
Qy	494 -----	493	Qy	753 YRKEDLKYQSSKOVGUDAWLK	774	
Db	601 KEVSGDGKIGEKGDLIVPKKIKUNEKGYSINIPELISLSEETADIVMTISAKGRKNEFK	660	Db	1681 YRKEDLRYQSSKOTGUDAWLK	1702	
Qy	494 -----	493	RESULT 2			
Db	661 GMURTLRMFGEENRIRTENRYLFLHEKGLIILKURKLYKOLYELAG	720	S67220			
Qy	494 -----	493	DNA-directed DNA polymerase (EC 2.7.7.7) - <i>Thermococcus</i> sp.			
Db	721 SVKYNGKREYLWFMNIEIKDFISYPPKKELEBWKIGTLNGFRNTCLKVDEDFGKLGY	780	C;Species: <i>Thermococcus</i> sp.			
Qy	494 -----	493	C;Accession: S67920			
Db	781 VSEGYAGAQNKNGGIGSYSSVKAQNVEAKFFGKVRVDRNCVTSISKMAYL	840	C;Date: 17-Feb-1998 #sequence_revision 13-Mar-1998			
Qy	494 -----	493	C;Accession: S67920			
Db	901 VFLINSLGTSVSKVIGFDSGVYRYVYINBLQFOTSREKNTYYSNLPIKEILRDVFKEFQ	960	C;Date: 17-Feb-1998 #text_change 18-Jun-1999			
Qy	494 -----	493	C;Accession: S67920			
Db	841 VMKCLCGALAENRIPSVLTSPEPVWSFLEAYFTGDIDHPSKRFLSTKSELANQL	900	C;Submitted to the EMBL Data Library, January 1996			
Qy	494 -----	493	C;Description: Cloning, expression and modulation of the 3'-5' exonuclease activity of			
Db	961 KNMTFKFKELVDSGKLNREKAKLLEFFINGDIVLDRVKSYKEKDYEYVYDLSYDENEN	1020	C;Accession: S67920			
Qy	494 -----	493	A;Molecule type: DNA			
Db	961 KNMTFKFKELVDSGKLNREKAKLLEFFINGDIVLDRVKSYKEKDYEYVYDLSYDENEN	1020	A;Residues: 1-775 <SOU>			
Qy	494 -----	493	A;Cross-references: GB:U471018; NID:91197451; PID:AAA88769.1; PID:91197452			
Db	961 KNMTFKFKELVDSGKLNREKAKLLEFFINGDIVLDRVKSYKEKDYEYVYDLSYDENEN	1020	A;Conservative: 99; Mismatches: 89; Indels: 3; Gaps: 2;			
Qy	494 -----	493	C;Superfamily: herpesvirus DNA-directed DNA polymerase			
Db	961 KNMTFKFKELVDSGKLNREKAKLLEFFINGDIVLDRVKSYKEKDYEYVYDLSYDENEN	1020	C;Keywords: DNA binding; nucleotidyltransferase			
Qy	494 -----	493	Qy 1 MIFTDTYITKDGKPIIRIFKKGNGEKFKEILDPHQPYTIALLKDSDAIDBKAKGERHGG 60			
Db	1 MIFTDTYITKDGKPIIRIFKKGNGEKFKEILDPHQPYTIALLKDSDAIDBKAKGERHGG 60	1	Db 1 MIFTDTYITKDGKPIIRIFKKGNGEKFKEILDPHQPYTIALLKDSDAIDBKAKGERHGG 60			
Qy	494 -----	493	Qy 61 KIVRYVDVAKVKKFLGRDVEVNLKIFEPHODVPAIRGKIREHPAVDIFYDIPFARY 120			
Db	61 KIVRYVDVAKVKKFLGRDVEVNLKIFEPHODVPAIRGKIREHPAVDIFYDIPFARY 120	1	Db 61 KIVRYVDVAKVKKFLGRDVEVNLKIFEPHODVPAIRGKIREHPAVDIFYDIPFARY 120			
Qy	494 -----	493	Qy 61 TTVVKVRAEKVQKFLGRPIEWKLRYFVNLKIFEPHODVPAIRGKIREHPAVDIFYDIPFARY 120			

Best Local Similarity	74.6%	Pred. No.	2.3e-154;	Indels	2;	Gaps	1;
Matches	576;	Conservative	100;	Mismatches	94;		
QY	1	MIFDDYKNDGKPLIRPKKENGEXKIELDPHQPIYALIKDSNADIEKAIGERHG	60				
Db	1	MILDVDTTEGKEVTRFKKENGKFKFKEHDTRTRPYTALRDDSKEEVKTRGKFRHG	60				
QY	61	KIVRVDAAVKKFKLGRDVEVKLIFEPQDYLKRAEKGUVTLIGRDEBHEPE	120				
Db	61	KIVIIVDVEKVEKFKLGRKTYKVKLYLHQDQPTIREPAVVDYEDIPPAKRY	120				
QY	121	LIDKGLIPMEGDEEIKLAFDIEPFYHEDFGKEIMISYADEEAEARVITWNIDLPY	180				
Db	121	LIDKGLIPMEGEESEKILAFDIELYHEGEEFKGKPIIMISYADEEAEARVITWNIDLPY	180				
QY	181	VDVVSNEREMIKRPFVQIVREKDPVLTINGDNFDLPLTYKRAEKGUVTLIGRDEBHEPE	240				
Db	181	VEVSSEREMIKRFLIRKDPDIVTYNGDSDFPFLAKRAEKGKLTIGRDS-E	238				
QY	241	PKIHMGDGDAVEAKGRTHEDFLPFVVRTRINLPYTLAIVEAFLGKTKSKLGAEETAAI	300				
Db	239	PKMQGIDMATAVEKGRHIDFLPFVVRTRINLPYTLAIVEAFLGKREKTYADEAKA	298				
QY	301	WETEBSMKKLAQYSMEDARATYELGKEFFPMEEAKLJGQSYDVSRSSTGMLVENVLL	360				
Db	299	WESEGNLRLTAKYSMEDARATYELGKEFFPMEEAKLJGQSYDVSRSSTGMLVENVLL	358				
QY	361	RVAYERNELLAPNPKDPEEYERRRITYLGGYKEPERGLMENITYLDFRCYLYPSITIWHN	420				
Db	359	RKAYERNEVAPNPKSEEYORRRLRESYGFGFVSEPEKGWENIVYDRLAYSITIWHN	418				
QY	421	VSPDILERECKNYDVAPIVGYKCFDPPOFIPSILGELITMPOEIKKMKATIDPIKK	480				
Db	419	VSPDILNLLEGCKNYDIAPOVGHFKCFDPPOFIPSILGELITMPOEIKKMKATIDPIKK	478				
QY	481	MLDYDRAVKLHANSYGYNGYPKARWYKSCAEGSVTAANGRHYETMKIIEEKGFRVY	540				
Db	479	LLDYRQAKILLANSFYGYYAKARWYCKECAEGSVTAANGRKYTELVNEKEEKGFRVY	538				
QY	541	YADTDGFYATIPEGEKPTIKKAKFLKYINSKLGFLLEBEYGRYLRGFFVAKKRYAVI	600				
Db	539	YIDTDGFYATIPEGESEKIKKALEVKVINSKLGFLLEBEYGRYLRGFFVAKKRYAVI	598				
QY	601	DEEGRITTRGLEYVRDWSLEAKETOAKYLEALKEDSYEKAELVQDVEETAKYQVPL	660				
Db	599	DEBKGKVTRGLEYVRDWSLEAKETOARVLETLKGDVBEAVR-LVKEVITQKLANYEF	658				
QY	661	EKLVISHBQITDSEKTAIGPVAVAKRLAAGKIGIIVRPGTISIYIYLRSGKTSDRVIL	720				
Db	659	EKLAIYQITDSEKTAIGPVAVAKRLAAGKIGIIVRPGTISIYIYLRSGKTSDRVIL	718				
QY	721	SEYDPKXKHIDDDYIENQULPAVRLLEANGYRKEDLKQYQSKQVGDawl	772				
Db	719	SEYDPKXKHIDDDYIENQULPAVRLLEANGYRKEDLKQYQSKQVGDawl	770				
QY	88593	DNA-directed DNA polymerase (EC 2.7.7.7) Vent, intein containing precursor - Pyrococcus	5				
Db	88593	DNA-directed DNA polymerase (EC 3.1.----) PI-I, DNA-directed DNA polymerase					
QY	88593	NC_Contents: DNA endonuclease					
Db	88593	C_Species: Pyrococcus sp.					
QY	88593	C_Date: 2-Aug-1996 #Sequence_revision 01-Nov-1996					
Db	88593	C_Accession: S68593					
QY	88593	R.Xu, M.Q.; Southworth, M.W.; Mersha, F.B.; Hornstra, L.J.; Perier, F.B.					
Db	88593	submitted to the EMBL Data Library August 1993					
QY	88593	A_Reference number: S68593					
Db	88593	A_Molecule type: DNA					
QY	88593	A_Cross-references: EMBL:U00707; NID:9436492; PID:AAA67130.1; PID:g825735					
Db	88593	R.Xu, M.Q.; Southworth, M.W.; Mersha, F.B.; Hornstra, L.J.; Perier, F.B.					
QY	88593	Cell 75, 1371-1377, 1993					
Db	88593	GMLRFLRWFGEERKRVTSRSLRHLNIGYTRLKGIGYD1DEGEGLEKRTLYERLUVDF					

A;Title:	In vitro protein splicing of purified precursor and the identification of a branched intermediate and its resolution						
R:Xu, M.Q.; Comb, D.G.; Paulus, H.; Noren, C.J.; Shao, Y.; Perier, F.B.							
A;Contents:	annotation						
R:Xu, M.Q.; Comb, D.G.; Paulus, H.; Noren, C.J.; Shao, Y.; Perier, F.B.							
A;Title:	Protein splicing: an analysis of the branched intermediate and its resolution						
A;Contents:	annotation						
R:Xu, M.Q.; Comb, D.G.; Paulus, H.; Noren, C.J.; Shao, Y.; Perier, F.B.							
A;Function:	<NCCLS>						
A;Description:	nucleotidyltransferase						
C;Function:	nucleotidyltransferase						
C;Function:	DNA-directed DNA polymerase Vent						
C;Function:	<ENDO>						
C;Function:	endonuclease						
C;Function:	hydrolyase						
C;Function:	endonuclease PI-PI						
C;Function:	hypothetical protein PH0202						
C;Superfamily:	hypothetical protein PH0202						
C;Keywords:	endonuclease; hydrolyase; nucleotidyltransferase; protein						
F:1-91/1312/Domain:	product: DNA-directed DNA polymerase Vent #status Predicted <MAT1>						
F:1-91/1312/Domain:	product: DNA-directed DNA Polymerase Vent extein 1 #status Predicted <XTL1>						
F:492/1029/Domain:	product: DNA endonuclease PI-I (pol Vent intein 1) #status Predicted <MAT2>						
F:1030/1312/Domain:	product: DNA-directed DNA Polymerase Vent extein 2 #status Predicted <XT2>						
F:491/1030/Cross-link:	peptide (Ala-Ser) #status experimental						
Query Match	72.9%						
Best Local Similarity	45.6%	Pred. No.	6.1e-143;				
Matches	597;	Conservative	86;	Mismatches	87;	Indels	539;
Gaps	3;						
QY	1	MIFDDYITTDGKPTIRIFKENGKFKIPEKQDPLTYKRAEKGUVTLIGRDEBHEPE	60				
Db	1	MILDAYITTEGKPTIRIFKENGKFKIPEKQDPLTYKRAEKGUVTLIGRDEBHEPE	60				
QY	61	KIVRVDAAVKKFKLGRDVEVKLIFEPQDYLKRAEKGUVTLIGRDEBHEPE	120				
Db	61	KIVRVAEAKVLLGKIGYTRKDPDITVYNGDSDFPFLAKRAEKGKLTIGRDS-E	120				
QY	121	LIDKGLIPMEGDEEIKLAFDIEPFYHEDFGKEIMISYADEEAEARVITWNIDLPY	180				
Db	121	LIDKGLIPMEGEESEKILAFDIELYHEGEEFKGKPIIMISYADEEAEARVITWNIDLPY	180				
QY	181	VDVVSNEREMIKRPFVQIVREKDPVLTINGDNFDLPLTYKRAEKGUVTLIGRDEBHEPE	240				
Db	181	VEVSSEREMIKRFLIRKDPDIVTYNGDSDFPFLAKRAEKGKLTIGRDS-E	238				
QY	241	PKIHMGDGDAVEAKGRTHEDFLPFVVRTRINLPYTLAIVEAFLGKTKSKLGAEETAAI	300				
Db	239	PKMQGIDMATAVEKGRHIDFLPFVVRTRINLPYTLAIVEAFLGKREKTYADEAKA	298				
QY	301	WETEBSMKKLAQYSMEDARATYELGKEFFPMEEAKLJGQSYDVSRSSTGMLVENVLL	360				
Db	299	WESEGNLRLTAKYSMEDARATYELGKEFFPMEEAKLJGQSYDVSRSSTGMLVENVLL	358				
QY	361	RVAYERNELLAPNPKDPEEYERRRITYLGGYKEPERGLMENITYLDFRCYLYPSITIWHN	420				
Db	359	RKAYERNEVAPNPKSEEYORRRLRESYGFGFVSEPEKGWENIVYDRLAYSITIWHN	418				
QY	421	VSPDILERECKNYDVAPIVGYKCFDPPOFIPSILGELITMPOEIKKMKATIDPIKK	480				
Db	419	VSPDILNLLEGCKNYDIAPOVGHFKCFDPPOFIPSILGELITMPOEIKKMKATIDPIKK	478				
QY	481	MLDYDRAVKLHANSYGYNGYPKARWYKSCAEGSVTAANGRHYETMKIIEEKGFRVY	540				
Db	479	LLDYRQAKILLANSFYGYYAKARWYCKECAEGSVTAANGRKYTELVNEKEEKGFRVY	538				
QY	541	YADTDGFYATIPEGEKPTIKKAKFLKYINSKLGFLLEBEYGRYLRGFFVAKKRYAVI	600				
Db	539	YIDTDGFYATIPEGESEKIKKALEVKVINSKLGFLLEBEYGRYLRGFFVAKKRYAVI	598				
QY	601	DEEGRITTRGLEYVRDWSLEAKETOAKYLEALKEDSYEKAELVQDVEETAKYQVPL	660				
Db	599	DEBKGKVTRGLEYVRDWSLEAKETOARVLETLKGDVBEAVR-LVKEVITQKLANYEF	658				
QY	661	EKLVISHBQITDSEKTAIGPVAVAKRLAAGKIGIIVRPGTISIYIYLRSGKTSDRVIL	720				
Db	659	EKLAIYQITDSEKTAIGPVAVAKRLAAGKIGIIVRPGTISIYIYLRSGKTSDRVIL	718				
QY	721	SEYDPKXKHIDDDYIENQULPAVRLLEANGYRKEDLKQYQSKQVGDawl	772				
Db	719	SEYDPKXKHIDDDYIENQULPAVRLLEANGYRKEDLKQYQSKQVGDawl	770				
QY	88593	DNA-directed DNA polymerase (EC 2.7.7.7) Vent, intein containing precursor - Pyrococcus	5				
Db	88593	DNA-directed DNA polymerase (EC 3.1.----) PI-I, DNA-directed DNA polymerase					
QY	88593	NC_Contents: DNA endonuclease					
Db	88593	C_Species: Pyrococcus sp.					
QY	88593	C_Date: 2-Aug-1996 #Sequence_revision 01-Nov-1996					
Db	88593	C_Accession: S68593					
QY	88593	R.Xu, M.Q.; Southworth, M.W.; Mersha, F.B.; Hornstra, L.J.; Perier, F.B.					
Db	88593	submitted to the EMBL Data Library August 1993					
QY	88593	A_Reference number: S68593					
Db	88593	A_Molecule type: DNA					
QY	88593	A_Cross-references: EMBL:U00707; NID:9436492; PID:AAA67130.1; PID:g825735					
Db	88593	R.Xu, M.Q.; Southworth, M.W.; Mersha, F.B.; Hornstra, L.J.; Perier, F.B.					
QY	88593	Cell 75, 1371-1377, 1993					
Db	88593	GMLRFLRWFGEERKRVTSRSLRHLNIGYTRLKGIGYD1DEGEGLEKRTLYERLUVDF					
QY	88593	MLDYRORAVK	490				
Db	88593	MLYQRALKILANSIPEWPLKNGKVKIFRGDFVDSGMLKANQGVKKTGDTVEVL	538				
QY	491	--IHA--	493				
Db	539	VAGTHAFSPDRSKRARMVAKFVAVRHRYSGNVRYVLSNGRKVSRKPTIEGHSLFVNRGNDL	598				
QY	494	--	493				
Db	494	VSPDTLREGCBREYDPAVEGHKCFDPGFTPSLKRLLDEROKRNMASKSOPIEKK	478				
QY	494	WETEBSMKKLAQYSMEDARATYELGKEFFPMEEAKLJGQSYDVSRSSTGMLVENVLL	360				
Db	299	WESEGNLRLTAKYSMEDARATYELGKEFFPMEEAKLJGQSYDVSRSSTGMLVENVLL	358				
QY	361	RVAYERNELLAPNPKDPEEYERRRITYLGGYKEPERGLMENITYLDFRCYLYPSITIWHN	420				
Db	359	RVAYERNEVAPNPKSEEYORRRLRESYGFGFVSEPEKGWENIVYDRLAYSITIWHN	418				
QY	421	VSPDTLREGCBREYDPAVEGHKCFDPGFTPSLKRLLDEROKRNMASKSOPIEKK	480				
Db	419	VSPDTLREGCBREYDPAVEGHKCFDPGFTPSLKRLLDEROKRNMASKSOPIEKK	478				
QY	481	MUDYRORAVK	490				
Db	479	MLYQRALKILANSIPEWPLKNGKVKIFRGDFVDSGMLKANQGVKKTGDTVEVL	538				
QY	491	--IHA--	493				
Db	539	VAGTHAFSPDRSKRARMVAKFVAVRHRYSGNVRYVLSNGRKVSRKPTIEGHSLFVNRGNDL	598				
QY	494	--	493				
Db	659	GMLRFLRWFGEERKRVTSRSLRHLNIGYTRLKGIGYD1DEGEGLEKRTLYERLUVDF	718				

Db	1138 GPHVAVAKRMLMARGIKVKPGMVGIVYIYLRGDPISKRAISIEEFDPRKHRYDAEYYIENQ	1197	Db	359 RAYERNELAPNPKDEKEKLARR- RQSTEGGYVYKEPERGLWENIVYLFRCHPADTKYVVK	417
Qy	740 VLPAVIRILEAFGYRKEDLKQSSKQVGLDAWLK 773	Qy	411	411	410
Db	1198 VLPAVIRILAKFGYKREDLWRQKTVQGLGAWIK 1231	Db	418 GKGININISVQEBDYVLDGMQRVRKVWEYDYKGELVNINGLKCPTNHLKPVVTKNERQ	477	411
Qy	RE5ULT 7	Qy	411	411	410
S71551	DNA-directed DNA polymerase (EC 2.7.7.7) KOD, intein containing precursor - Pyrococcus s	Db	478 TRIDSLAKSFLTKVKGERIITTPFLFYIGRATSENIPPEEVLKGEGLAILLAGETLRLK	537	411
N; Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA endonuclease (EC 3.1.-.-) PI-II; DNA	Qy	411	411	410	411
C; Species: Pyrococcus sp.	Db	538 DVEFDSSRKKRER-SHQVRVEITIGKDEEEFRDRITYFERLFGITTPS-SEKKGTAIVTL	597	411	411
A; Variety: strain KOD1	Qy	598 KVAKKNVYLKVKEIMDNIESLHAPSVLRGFFEGDSVNRSSVATOGTKNEWKIKLVS	657	411	410
C; Accession: S71551	Db	658 KLLSOLGIPHOTYTYQOENGKDRSRYILETGKDGGLFQTLIGFISERKNALLNKAIS	717	411	410
R; Kakihara, H.; Takagi, M.; Imanaka, T.	Qy	718 QREHNLLENGFYRISSEFNSTTEYEGKYDILTLEGTPTYFANGILTHNSLYPSITIEN	777	411	410
submitted to the EMBL Data Library, March 1994	Db	778 VSPDTINREGKEYDVAQYGHRFKDQFQFIPSLGDLLEERQKIKKOMKATDPIERK	837	411	410
A; Description: Cloning and analysis of the DNA polymerase gene from a new hyperthermophil	Qy	838 LDYDORAIIKILANSILPEEMLPVLEGEVHVRIGELIDRMMEENAGKVKREGETEVLE	897	411	410
A; Reference number: S71551	Db	898 VSGLIEVPSFNRRRTNAELKRVKALIHDYGGKVITRLKGRRKIKITSGHSLFSRNGEL	957	411	410
A; Molecule type: DNA	Qy	944 - - - - -	494	493	493
A; Residues: 1-1670 <KAK>	Db	944 - - - - -	494	493	493
A; Cross-references: EMBL:D29671	Qy	944 - - - - -	494	493	493
C; Function: <NUCL>	Db	944 - - - - -	494	493	493
A; Description: as DNA directed DNA polymerase, catalyzes the polymerization of DNA at th	Qy	944 - - - - -	494	493	493
A; Note: DNA-directed DNA polymerase KOD	Db	1018 GMLRITLRWIFGEEKRRTARRYLRHL-EDLGTYVRLKIGYEVLDWDSLNTERLYEALVEN	1071	493	493
C; Function: <EN1>	Qy	1018 GMLRITLRWIFGEEKRRTARRYLRHL-EDLGTYVRLKIGYEVLDWDSLNTERLYEALVEN	1071	493	493
A; Description: as DNA endonuclease PI-I, catalyzes the hydrolysis of internal phosphodi	Db	1078 VRYNGNKRKLYVEFNSIRAVGIMPLKELKEWKIGTLNGFMRKLIEVEDSLAKLIGYYY	1137	493	493
A; Note: DNA endonuclease PI-PspI	Qy	1078 VRYNGNKRKLYVEFNSIRAVGIMPLKELKEWKIGTLNGFMRKLIEVEDSLAKLIGYYY	1137	493	493
C; Function: <EN2>	Db	1138 SEGYARKORNPKNGWSVLYNEDFPEVLDMDMERLASRFFGKVYGRNRYVEIPKIGYLL	1197	493	493
A; Description: as DNA endonuclease PI-II, catalyzes the hydrolysis of internal phosphodi	Qy	1138 SEGYARKORNPKNGWSVLYNEDFPEVLDMDMERLASRFFGKVYGRNRYVEIPKIGYLL	1197	493	493
C; Superfamily: DNA-directed DNA polymerase KOD	Db	1194 - - - - -	494	493	493
Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing	Qy	1194 - - - - -	494	493	493
F71-406; Domain: DNA endonuclease; hydrolase; nucleotidyltransferase; protein splicing	Db	1258 FENMCGYLAENKRIPEFVFTSPKGVRILAFLFGYSSAMATSPEQETQALNERALAQVL	1257	493	493
F71-407; 767-851/Domain: DNA-directed DNA polymerase KOD extein 1 #status Predicted <XT1>	Qy	1258 FENMCGYLAENKRIPEFVFTSPKGVRILAFLFGYSSAMATSPEQETQALNERALAQVL	1257	493	493
F71-407-766/Product: DNA endonuclease; hydrolase; nucleotidyltransferase; protein splicing	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
F71-767-851/Domain: DNA-directed DNA polymerase KOD extein 1 #status Predicted <XT1>	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
F71-852-1387/Product: DNA endonuclease KOD extein 2 #status Predicted <XT2>	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
F71-852-1387/Product: DNA endonuclease KOD extein 2 #status Predicted <XT2>	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
F71-1388-1670/Domain: DNA-directed DNA polymerase KOD extein 3 #status Predicted <XT3>	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
F71-406-767/Cross-link: Peptide (Arg-Ser) #status Predicted	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
F71-851-1388/Cross-link: Peptide (Asn-Ser) #status Predicted	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
Query Match	68.7%: Score 2764.5; DB 2; Length 1670;	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493
Best Local Similarity	35.9%: Pred. No. 4 .Se-13;	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493
Matches 599; Conservative	83; Mismatches 88; Indels 899; Gaps 4;	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493
1 MIFDTDITKDGPIIPIKFKKENGEFKTKEFLDPHQPYIALLKDDSALEDEIKIGERHG 60	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
1 MILDTDYTEDGKEVIRIIFKKEGEFKYEYDRTEPYALLKDDSALEEVKTKTARHG 60	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
61 KIVYVVDAVKVKQKFLGRDVEVWKLIFPHQDYLALRGTREPAVIDYDIPFARRY 120	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
61 TVVTVKVRKEVKQVKFLGRDVEVWKLIFPHQDPAIRKIREGAVIDYDIPFARRY 120	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
1221 LIDGLLPMGDEEIKLMAFDIEFYHBDPEKFLGRDVEVWKLIFPHQDYLALRGTREPAVIDYDIPFARRY 180	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
1221 LIDGLLPMGDEEIKLMAFDIQFLYHEGBEPARGPILMISYADEGGRVITKVNIDPY 180	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
181 VDVSNEREMVKRQIVREKDPVLIYTYGDNFLDPLVLIKRAEKLGTLLGRDKERPE 240	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
181 VDVSNEREMVKRQIVREKDPVLIYTYGDNFLDPLVLIKRAEKLGTLLGRDKERPE 240	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
241 PKIHMGDPSAVEIKGRTHPDLPVYRVRTINLPYTLAYEAVLGTKSKLGAEEIAAI 300	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
239 PKIQNGDPSAVEIKGRTHPDLPVYRVRTINLPYTLAYEAVLGTKSKLGAEEIAAI 298	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
301 WETESEMKKLAQSMEDARATYEGKEFPEMEEIAKLIQGOSVDRSRSSTGNLIVENVIL 360	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
299 WETGNNLERYARYSNEKDYKTYELGKPSLQSLGQSLIDVRSRSSTGNLIVENVIL 358	Qy	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
361 RVAYERNELAPNPKDEKEKLARR- RQSTEGGYVYKEPERGLWENIVYLFRCHPADTKYVVK 417	Db	1318 VSPQTERKVEDGRDPEKAQRLSLWIEBGDVYDLYRVEVDYDGTYYDLDVENDENFL	1371	493	493
Qy	RESULTS	Qy	411	411	410
S71551	DNA-directed DNA polymerase (EC 2.7.7.7) KOD, intein containing precursor - Pyrococcus s	Db	478 TRIDSLAKSFLTKVKGERIITTPFLFYIGRATSENIPPEEVLKGEGLAILLAGETLRLK	537	411
N; Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA endonuclease (EC 3.1.-.-) PI-II; DNA	Qy	478 TRIDSLAKSFLTKVKGERIITTPFLFYIGRATSENIPPEEVLKGEGLAILLAGETLRLK	537	411	410
C; Species: Pyrococcus sp.	Db	538 DVEFDSSRKKRER-SHQVRVEITIGKDEEEFRDRITYFERLFGITTPS-SEKKGTAIVTL	597	411	410
A; Variety: strain KOD1	Qy	598 KVAKKNVYLKVKEIMDNIESLHAPSVLRGFFEGDSVNRSSVATOGTKNEWKIKLVS	657	411	410
C; Accession: S71551	Db	658 KLLSOLGIPHOTYTYQOENGKDRSRYILETGKDGGLFQTLIGFISERKNALLNKAIS	717	411	410
R; Kakihara, H.; Takagi, M.; Imanaka, T.	Qy	718 QREHNLLENGFYRISSEFNSTTEYEGKYDILTLEGTPTYFANGILTHNSLYPSITIEN	777	411	410
submitted to the EMBL Data Library, March 1994	Db	778 VSPDTINREGKEYDVAQYGHRFKDQFQFIPSLGDLLEERQKIKKOMKATDPIERK	837	411	410
A; Description: Cloning and analysis of the DNA polymerase gene from a new hyperthermophil	Qy	838 LDYDORAIIKILANSILPEEMLPVLEGEVHVRIGELIDRMMEENAGKVKREGETEVLE	897	411	410
A; Reference number: S71551	Db	898 VSGLIEVPSFNRRRTNAELKRVKALIHDYGGKVITRLKGRRKIKITSGHSLFSRNGEL	957	411	410
A; Molecule type: DNA	Qy	944 - - - - -	494	493	493
A; Residues: 1-1670 <KAK>	Db	944 - - - - -	494	493	493
A; Cross-references: EMBL:D29671	Qy	944 - - - - -	494	493	493
C; Function: <NUCL>	Db	944 - - - - -	494	493	493
A; Description: as DNA directed DNA polymerase, catalyzes the polymerization of DNA at th	Qy	944 - - - - -	494	493	493
A; Note: DNA-directed DNA polymerase KOD	Db	944 - - - - -	494	493	493
C; Function: <EN1>	Qy	944 - - - - -	494	493	493
A; Description: as DNA endonuclease PI-I, catalyzes the hydrolysis of internal phosphodi	Db	944 - - - - -	494	493	493
A; Note: DNA endonuclease PI-PspI	Qy	944 - - - - -	494	493	493
C; Function: <EN2>	Db	944 - - - - -	494	493	493
A; Description: as DNA endonuclease PI-II, catalyzes the hydrolysis of internal phosphodi	Qy	944 - - - - -	494	493	493
C; Superfamily: DNA-directed DNA polymerase KOD	Db	944 - - - - -	494	493	493
Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing	Qy	944 - - - - -	494	493	493
F71-406; Domain: DNA endonuclease; hydrolase; nucleotidyltransferase; protein splicing	Db	944 - - - - -	494	493	493
F71-407; 767-851/Domain: DNA-directed DNA polymerase KOD extein 1 #status Predicted <XT1>	Qy	944 - - - - -	494	493	493
F71-767-851/Domain: DNA-directed DNA polymerase KOD extein 1 #status Predicted <XT1>	Db	944 - - - - -	494	493	493
F71-852-1387/Domain: DNA-directed DNA polymerase KOD extein 2 #status Predicted <XT2>	Qy	944 - - - - -	494	493	493
F71-852-1387/Domain: DNA-directed DNA polymerase KOD extein 2 #status Predicted <XT2>	Db	944 - - - - -	494	493	493
F71-1388-1670/Domain: DNA-directed DNA polymerase KOD extein 3 #status Predicted <XT3>	Qy	944 - - - - -	494	493	493
F71-406-767/Cross-link: Peptide (Arg-Ser) #status Predicted	Db	944 - - - - -	494	493	493
F71-851-1388/Cross-link: Peptide (Asn-Ser) #status Predicted	Qy	944 - - - - -	494	493	493
Query Match	68.7%: Score 2764.5; DB 2; Length 1670;	Db	944 - - - - -	494	493
Best Local Similarity	35.9%: Pred. No. 4 .Se-13;	Qy	944 - - - - -	494	493
Matches 599; Conservative	83; Mismatches 88; Indels 899; Gaps 4;	Db	944 - - - - -	494	493
1 MIFDTDITKDGPIIPIKFKKENGEFKTKEFLDPHQPYIALLKDDSALEDEIKIGERHG 60	Db	944 - - - - -	494	493	493
1 MILDTDYTEDGKEVIRIIFKKEGEFKYEYDRTEPYALLKDDSALEEVKTKTARHG 60	Qy	944 - - - - -	494	493	493
61 KIVYVVDAVKVKQKFLGRDVEVWKLIFPHQDYLALRGTREPAVIDYDIPFARRY 120	Db	944 - - - - -	494	493	493
61 TVVTVKVRKEVKQVKFLGRDVEVWKLIFPHQDPAIRKIREGAVIDYDIPFARRY 120	Qy	944 - - - - -	494	493	493
1221 LIDGLLPMGDEEIKLMAFDIEFYHBDPEKFLGRDVEVWKLIFPHQDYLALRGTREPAVIDYDIPFARRY 180	Db	944 - - - - -	494	493	493
1221 LIDGLLPMGDEEIKLMAFDIQFLYHEGBEPARGPILMISYADEGGRVITKVNIDPY 180	Qy	944 - - - - -	494	493	493
181 VDVSNEREMVKRQIVREKDPVLIYTYGDNFLDPLVLIKRAEKLGTLLGRDKERPE 240	Db	944 - - - - -	494	493	493
181 VDVSNEREMVKRQIVREKDPVLIYTYGDNFLDPLVLIKRAEKLGTLLGRDKERPE 240	Qy	944 - - - - -	494	493	493
241 PKIHMGDPSAVEIKGRTHPDLPVYRVRTINLPYTLAYEAVLGTKSKLGAEEIAAI 300	Db	944 - - - - -	494	493	493
239 PKIQNGDPSAVEIKGRTHPDLPVYRVRTINLPYTLAYEAVLGTKSKLGAEEIAAI 298	Qy	944 - - - - -	494	493	493
301 WETESEMKKLAQSMEDARATYEGKEFPEMEEIAKLIQGOSVDRSRSSTGNLIVENVIL 360	Db	944 - - - - -	494	493	493
299 WETGNNLERYARYSNEKDYKTYELGKPSLQSLGQSLIDVRSRSSTGNLIVENVIL 358	Qy	944 - - - - -	494	493	493
361 RVAYERNELAPNPKDEKEKLARR- RQSTEGGYVYKEPERGLWENIVYLFRCHPADTKYVVK 417	Db	944 - - - - -	494	493	493

Qy	605 RITTRGLEVVERDWEIAKE TOAKVLEPAIILKEDSV EKAVEIIVKDVYEEIAKYQVPLEKLV 664	Db	453 FFKRILMLIEKRREKLVELKNLSPESSEYKLLDIKQOTLKVLTNSFYGGMGNLARWYC 512
Db	1498 KITRGLLIVRRDWEIAKE TQARVLEPAIILKEDSV EKAVEIIVKDVYEEIAKYQVPLEKLV 1557	Qy	510 KCEADESVIATWGRHYIEMTIKEBEKFGPKRYTADTDGYATIPGEKPTIKKAKBFKLY 569
Qy	665 IHEQITKULSEYEKAIGPVIAIAKRLAAGKIKYVRPGTISI SY1VLRGSGKISDRVILLSEYD 724	Db	513 HPCHEATTAIWGRHPTIRTBAK-TAESMGKVYQGTDTSFPTVYKAGMTEDVDR--JIDK 567
Db	1558 IHEQITRLLKDYKATGPVAVAKRLAARGVKRPGTISI SY1VLRGSGRIGDRAIPFDEFD 1617	Qy	570 INSKPLGILLEBEYEGFYLRLGFFYAKKRAYTIDBEGRITTRGLEVVRDWEIAKE TOAKV 629
Qy	725 PKCHKYDDYYTENQVLPAVLRLAEGYRKEKDLYQSKVQLDMLK 773	Db	568 LHEEP- -IQIEVEDEYSSAIIFFYKERYAGLTDGRLVKGLEVRGDWCELAKKVQREV 625
Db	1618 PTXHKYDDYYTENQVLPAVLRLAEGYRKEKDLYQSKVQLDMLK 1666	Qy	630 LEAIKEDSVERKAVEIIVKDVYERIAKCYOPLKLYTHEQITDSEYKAIAGPHVIAKRL 689
RESULT 8			
A9312	DNA polymerase B1 (polB) homolog - Archaeoglobus fulgidus	Qy	630 LEAIKEDSVERKAVEIIVKDVYERIAKCYOPLKLYTHEQITDSEYKAIAGPHVIAKRL 689
CSpecies: Archaeoglobus fulgidus			
CDate: 05-Dec-1997 #text_change 22-Oct-1999			
CAccession: A69132			
R.Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Giedek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; McDonald, L.			
Nature 390, 364-370, 1997			
A; Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.			
A; Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon ARRAference number: A69250; MUI:9804934; PMID:9389475			
A; Accession: A69132			
A; Status: Preliminary; nucleic acid sequence not shown; translation not shown			
A; Molecule type: DNA			
A; Residues: 1-781 <KLE>			
A; Cross-references: GE:AE0001070; GB:AE000782; NID:92689393; PIDN:AA80741.1; PID:9265013			
C; Superfamily: herpesvirus DNA-directed DNA Polymerase			
Query Match Score 1301; DB 2; Length 781;			
Best Local Similarity 37.4%; Pred. No. 2.3e-59; Matches 305; Conservative 155; Mismatches 266; Indels 90; Gaps 21;			
Qy	2 IFDIDTYTIDKGKPIIIRFKKENGEFKTLEDPHDFQPYIYALKDDSAIDEKAIGKERHCK 61	Db	8 LIDADYETIGKAVVRWICDKDQGIVAYDNEPDYFVIGYD 64
Db	8 LIDADYETIGKAVVRWICDKDQGIVAYDNEPDYFVIGYD 64	Qy	62 IYVVDATVKVKKFKLGRDVEVKLIFPKYKTEKVERI PFKRHPAVIDYEDIPPAKYL 121
Qy	63 VIKLKSFKFAQKLTGGRVEGTYIYAHPOHVKDLSQFG--DVERADIPPAKYL 121	Db	122 IDKGK---IPEGDE---ELKLMADPDIETPYHEG-DEFG 153
Db	122 IDKGK---IPEGDE---ELKLMADPDIETPYHEG-DEFG 153	Qy	122 IDKGK---IPEGDE---ELKLMADPDIETPYHEG-DEFG 153
Qy	154 KGBJIMSYADEBEARVITWNIDLPVYDVSNEREMI KRFVQIVREKDPVILTYNGDN 213	Db	182 KDPIVIVSKVNTDDDEI-----LGDERKIKSDFVPLKSYDPLIVCYNQDA 230
Db	182 KDPIVIVSKVNTDDDEI-----LGDERKIKSDFVPLKSYDPLIVCYNQDA 230	Qy	214 FDPYLLKRAEKLGVTLGRDKHEPEPKIHMRGDSFAVEIKGRTHFDLFPVVRTINLP 273
Db	231 FDPYLLKRAEWNIPIDVGRIGSN---VVFRGG---RPKTPGRINVDLIDAMRISDX 284	Qy	274 TYLEA7EAVLGIKTK-SKLGAEEIAIWEYBESMKLQAQSMDAREAYELGKEFFPME 332
Db	265 IKKLEN7AFLGKTKIADEIAKDIYWSRGEK-EKVLNTARQDANTYLIKEELLPH 343	Db	333 AELAKLIGOSVNDSYRSSTGNLVEWYLLRVAYERNELAPNKPDEBEYRRRLRTTYLGGY 392
Qy	344 YELSKMIRLPVDDVTRGRKQYDWWLSEAKIGIAPNPPEHAE----SYEGAFV 396	Db	393 KEBERGLKENTYLDPROLYPSITVHNNSPDTLREGCKN- -YDVAIPTIGYKPKDFPG 450
Db	397 LEERGIGHENVQALDFASMYRSPMIAFNISPDTY---GCRDCCYE APEVGHFKPSPDG 452	Qy	56 GBRHKGIVRUVDAVKKKK---FLGRDVEVWKLIFEPHDVPAIPLRKREHPAVDIBYDI 114
Qy	451 FIPSILGELITINRQEIKKRMK-ATIDPIEKNMIDYQRAVKLHANSYYGNGYPKARWYS 509	Db	69 ND---LKEVENIEVKILKILKEREVKIKIATHPKVPLR-KXKECEVKEIHD1 123

Qy	115 PFAKRYLIDKGILPM-----EGDE-----ELKLMADFIEDTFYHEGDEFGKGE-IMMIS 161	4.94	-----	4.93
Db	124 PFAKRYLIDNEIIPMTWYDFENKKPVSVBIPKLUKVADMEVNRDTEPNPBDPLMAS 183	4.94	-----	4.93
Qy	162 YADEESEARVITWNIDLIPYDVSNERMIKRETVQIVREKDPVILYNGDNFDPLIK 221	4.94	-----	4.93
Db	184 FWDENGKIVITYGEFNPNTEWVNKEBLIKKJLIEKEY--DVIYTNGDNFDPLKA 241	4.94	-----	4.93
Qy	222 RAEKLGVTLILGRDKEHPEPKHMRGDSFAVEIKGRHFDLFFPVYRRTINLPTTYLEAVY 281	4.94	-----	4.93
Db	242 RAKTYGIDNLNGKDEB--ELKIKRGGMMEYRSYIPIGRVIDLPIYISRRLKLTKYTEDVV 299	4.94	-----	4.93
Qy	282 BAVLGKTKSKLGAEBIAIAWETESEMMKLAQYSMEDABATYELGKEFPMEAEFLAKLGQ 341	4.94	-----	4.93
Db	300 YNLFGTEKLIKPIHITKIVDWNND--KTLIEYESQDACYTKYQKGYKFPELEYMFSRIVNQ 357	4.94	-----	4.93
Qy	342 SVWYRSSTGNLNEWYLLRVAYEENELAPNKPDEEEVRRRLATTILOGYTKEPERGLWE 401	4.94	-----	4.93
Db	358 TPEEITRMSSQMVVEYLMKRAFTXNMIVENPKPDEEEVRRVLTTEYECGYTKEPEKAFE 417	4.94	-----	4.93
Qy	402 NITYLDFRC-----	4.94	-----	4.93
Db	418 DIISMDPFRCHPKGTIVYKVGKGTIVNIEDYKEEGNYLGDGQWQKVKYKVEYEGELINN 477	4.94	-----	4.93
Qy	411 -----	4.94	-----	4.93
Db	478 GLKCTPNHK1PLRYKTKKKTNNDYLVRDYAKSLLTKFGECKLICKDFETIGNYK 537	4.94	-----	4.93
Qy	538 YINMDDBFLKSELIGLLAEGHLLRDTBYFDSRGGKRISHQYRBEITVNEDEDKFI 597	4.94	-----	4.93
Db	411 -----	4.94	-----	4.93
Qy	598 EKIKYIFKLFNELYVRRKGKTRATLGCAKDKIYKLEBILNKNEYKLTPLNATLRRGFFE 657	4.94	-----	4.93
Db	658 GDGYNTVRAVVNNQCTNNYDKIKFIASTLDRLGKXSYTYSYERGRKLYKVIIF 717	4.94	-----	4.93
Qy	411 -----	4.94	-----	4.93
Db	718 SKGLIKFSLISFISRKRNNNLNEITIROTLYKIGDGYFYDLDVCSLESYKGEVYL 777	4.94	-----	4.93
Qy	411 -----	4.94	-----	4.93
Db	778 TLEGRPYYFANGLTHNSLYPSIIISNISPTLDCECK--DVSEKLGHWFCKKKGEL 835	4.94	-----	4.93
Qy	452 IPSILGELITMRQEIKKMK--ATIDPI--ERKMLDYRORAVKLA-----493	4.94	-----	4.93
Db	836 IPIKTLRNLIERNIKERMKMAEIGINEEYNLDEQKSILKILANSILPDEYTIEE 895	4.94	-----	4.93
Qy	494 -----	4.94	-----	4.93
Db	896 DGIKVVKIGEYIDDLMRKHKDKIKKFGSISIELETNKIKTFSDEKIKKVKALIRH 955	4.94	-----	4.93
Qy	494 -----	4.94	-----	4.93
Db	956 PYFGKAYKIKLRSRGTIKVTRGHSLFVYENGKIVEVKDDYRFGDLIVVKKKLTCYDKEV 1015	4.94	-----	4.93
Qy	1016 VNIPIKRJINADEEEIKDOLVITRHKDIAFFVFLKKTLEDIENNKLVIFDDCILYKELG 1075	4.94	-----	4.93
Qy	1076 LIDNIKKINKVYDILDEEKPAKYYKCFDTVIEHGNFKKGRCNICKYIKRDYIANIPD 1135	4.94	-----	4.93
Db	1136 KEFEDCEIGAYSGKINALLKLDEKLAFLGFFVTRGRULKKKGETVYEISVYKSUPEY 1195	4.94	-----	4.93
Qy	1196 QKEIAETFKEVFGAGSMYKDKVIMDKNKIVLYKLYKIDLFLNLYGIFTPTVKGKGYKLTLNPKY 1255	4.94	-----	4.93
Db	1256 SFLDFGLKAKKNKSHRGTSMAKDEKYLQNLMLFLNLYGIFTPTVKGKGYKLTLNPKY 1315	4.94	-----	4.93
Qy	1316 GTVKOLMDEVKELEAFEYSGYYDLSVEDNENFLVNTYAHNSVYGTAFPARFYSRE 1375	4.94	-----	4.93
Db	1376 CAEVTYLGRKYIETTYKE AEKCFKVLXIDTGFYAA-IWKEKSKBLLKAMEFVEY 1433	4.94	-----	4.93
Qy	512 CAESVIAWGRHYIETNTIKEEKGKFLYIADTGFYATIGEK--PETIKKKAKFELKY 569	4.94	-----	4.93
Db	570 INSKPLGLBLBEYBGRYLRGFVAKKRYAVI'DEGRITRTRGLEVRRDWSEIAKTOAKV 629	4.94	-----	4.93
Qy	1434 INSKPLGTMLEFFGYFKRGIIVTUKKRYALIDENGRTVKGLEFVRRDNMIAKTTQREV 1493	4.94	-----	4.93
Db	630 LEAIIKEDSVKAEEVTKDVEEIAKRYQVPLEKLVTHEQTTKDLSSEYAKGPHVIAKRL 689	4.94	-----	4.93
Qy	1494 LEALVVEGSTEKAKLQIYDVKDLEKKTKEKDLIYTQLTKDPEKXTTAPHYIAKXL 1553	4.94	-----	4.93
Db	690 AAKGIIKVPGTISIVLRLGSGKIDDRVILSEYPKGKHYDPEPYTENQVLPALVRLIE 749	4.94	-----	4.93
Qy	1554 NREGRRIKVGDGIIIVTGVGKTSISERAKLPEEVD--IDDDIVNTYIDNQILPPVLRINE 1611	4.94	-----	4.93
Db	750 AFGYKEDLKYQSOXKQVGLDAWK 773	4.94	-----	4.93
Db	1612 AVGSKNEILKEGA-GLTLDKRFK 1634	4.94	-----	4.93
RESULTS 10				
C69028	DNA-dependent DNA polymerase family B (PolB1) - Methanobacterium thermoautotrophicum	4.94	-----	4.93
C;Species: Methanobacterium thermoautotrophicum	5.00	-----	4.93	
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999	5.00	-----	4.93	
C;Accession: C69028	5.00	-----	4.93	
R;Smith, D.R.; Boucette-Stramm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldridge, T.; Qiu, D.; Spadato, R.; Vicaire, R.; Wang, Y.; Wierbowski, J.; Gibson, R.; Jiwani, T.; Jiwani, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.; J. Bacteriol. 179, 7135-7155, 1997	5.00	-----	4.93	
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fun	5.00	-----	4.93	
A;Accession number: A69000; PMID: 9803754; PMID: 931463	5.00	-----	4.93	
A;Status: preliminary; nucleic acid sequence not shown; translation not shown	5.00	-----	4.93	
A;Molecule type: DNA	5.00	-----	4.93	
A;Cross-references: GB:AE000888; GB:AE000666; NID:92622304; PIDN:AA85697.1; PID:92622	5.00	-----	4.93	
A;Experimental source: strain Delta H	5.00	-----	4.93	
C;Genetics:	5.00	-----	4.93	
A;Gene: MTH1208	5.00	-----	4.93	
Query Match	30.0%	Score 1209.5;	DB 2;	Length 586;
Best Local Similarity	42.9%	Pred. No. 7.7e-55;		
Matches 255;	Conservative 107;	Mismatches 176;	Indels 57;	Gaps 10;
Qy	1 MIFPTDITYTKDGPPIIRIFKKB--NGFHKIEUDPHFQPYIYALLKD-DSATDEIKAGK 56	5.00	-----	4.93
Db	7 VLLDDIVTVDEVFVIRLEFGDKGSNEPIIAHRSRSPYIAKPTDOLDECIRLEEL 66	5.00	-----	4.93
Qy	57 ERHSGKIVRYVDAYKVK-KFLGRDVEYKLLIFEPHQDYPALRKIRBPAVIDYEDIP 115	5.00	-----	4.93
Db	67 EK--	5.00	-----	4.93
Qy	116 FAKRYLIDKGLIPB-----GDEELKLMAFDI 142	5.00	-----	4.93
Db	117 FYRFLIDDSIVPMEELFQGVENDSAPSVTTDVTVEVTRVQSTGAGHLDLFDI 176	5.00	-----	4.93
Qy	143 ETFHSEG-DEFGKGEIIMISYA-DEEARVITMKNIDLPYDVYNSNREMIKRFVQIVRE 200	5.00	-----	4.93
Db	177 EVRNPHGMPPEKEOBIVMGVAGNGYYESVSTAGDHLDLFDVEVDEELLERFAEIVD 236	5.00	-----	4.93

QY	320	ATYBEGKEFFPWEAELAKLIGQSVWVDVSRSSTGNNLVEWLLRVAYERNELAPNKPDBEEY	379
Db	348	SAVLLGEFIPPGIEURTRISGEPLDQSLMASGYHRVWLLREREAYCNELIPK-BBERY	406
QY	360	RRLRLTYLGGVYKEPERGLWNNTYDFFRLYPSLIVTHNVSPPDTLREGCQNYDVAPI	439
Db	407	E----SYEGGLVITSP1PGIHEEYVYDFFSSNTPSMIKYNGPDTLYKGCECNYWSP-	460
QY	440	VGYKPCKDFFGFTPSLIGELITMROEIKKOKKATIDPIEKGNDYQRAVKHANSYYGY	499
Db	461	VGHKFRKEPPGLYKNTVLEKLIQERKEVYKCLMKTIDDEYDKVLDARQRAKYMANYGY	520
QY	500	MGYPKAWNSKECAESVTAWGHYIENITKELEBEKSFKVLYADTDSFYATIPLGEKPBETI	559
Db	521	MGWLGRWVSKREGAIAVTAWGRQIISDSAKIAKEK-GFTVYQGDTDSIFVKGGD----	574
QY	560	KRKAKBPUKYNISKPLGSDLELEYEGFYLGRGFFVA-KKRYAVTDEEGRITTRGLEVYRWD	618
Db	575	--INSLITEISSKFG--LEIKDKYKRVFPTENKRYAGITEDKIDIVGFEARFGDW	629
QY	619	SEIAKETQAKYLEAIIKEDSYEKAVEIVKDVEEELAKEYQVPLEKLYTHEQITKDLSEYKA	678
Db	630	CDLAKQVOTNVNIELLXSGKVEDAIKIVKSUTFDRYNFRIEDLILWKTIDRNLDYDV	689
QY	679	IGPHVATAKRLAAKGKIVKRPGTISTIVLRSGGSKISDRV--ILLSEYDPPKCHKYPDYY	735
Db	690	TAPHVIAKKAAGLIVSKSKVKGIVVKGSKISDKAEPYFLVKE---KNNKIDVBYY	745
QY	736	TENQVLPAVRLILEAFGYRKEDLK	759
Db	746	IDKQITPVALRILEFGPKVKESSLK	769

A; Accession: S75407
A; Status: nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A; Residues: 1-764 <SEN>
A; Cross-references: EMBL:Y08257; NID:91707772; PID:91707813
A; Experimental source: strain P2
A; Note: the nucleotide sequence was submitted to the EMBL Data Library
C; Superfamily: herpesvirus DNA-directed DNA Polymerase
C; Keywords: nucleotidyltransferase

1 information content of an archaeal genome: 1

Result No.	Query	Score	Match	Length	DB	ID	Description
1	3446.5	85.6	1829	1	DPOL_THEST		O33845 thermococcus
2	3235.6	83.4	1702	1	DPOL_THELI		P30317 thermococcus
3	3253.5	80.8	773	1	DPOL_THEGO		P56889 thermococcus
4	3224.5	79.6	775	1	DPOL_THEB9		Q56316 thermococcus
5	3114.5	78.6	771	1	DPOL_PYRAB		P77916 pyrococcus
6	3131.5	78.3	775	1	DPOL_PYRAB		P80061 pyrococcus
7	2924.5	72.9	1312	1	DPOL_PRSID		Q51314 pyrococcus
8	2886	71.7	1235	1	DPOL_PYRAB		Q59610 pyrococcus
9	2809	69.8	1523	1	DPOL_PYRAB		P74918 thermococcus
10	2782	69.1	1671	1	DPOL_PYRKO		P77933 pyrococcus
11	2795.5	68.8	1699	1	DPOL_THEGB		Q9hb84 thermococcus
12	2535.5	63.0	1668	1	DPOL_THEHY		P52025 methanococcus
13	1339.2	34.6	824	1	DPOL_NETVO		P29753 archaeoglob
14	1301	32.3	781	1	DPOL_ARCTA		Q58235 methanococcus
15	1292	32.1	1634	1	DPOL_METJA		O27276 methanobacter
16	1209.5	30.0	586	1	DPOL_METH		P45346 aeropyrum
17	1138	28.3	784	1	DPOL_AERPE		P09462 oryza sativa
18	697.5	17.3	1105	1	DPOL_ORYSA		P30316 schizosaccharomyces
19	693.5	17.2	1088	1	DPOL_SOYBN		P54158 drosophila
20	621.5	17.2	763	1	DPOL_STLISH		P15436 saccharomyces
21	676.5	16.8	764	1	DPOL_SULSO		P54747 rattus norvegicus
22	676.5	16.8	1081	1	DPOL_ARATH		Q91vm7 arabidopsis
23	657.5	16.3	1038	1	DPOL_CANAL		P45389 candida albicans
24	657.5	16.3	1086	1	DPOL_SCHPO		P30316 schizosaccharomyces
25	624.5	15.5	1092	1	DPOL_DROME		P95970 sulfobolus
26	621.5	15.4	1097	1	DPOL_YEAST		P15436 saccharomyces
27	619.5	15.4	1103	1	DPOL_RAT		P54747 rattus norvegicus
28	616.5	15.3	872	1	DPOL_SULCH		O50307 sulfurisphaera
29	616.5	15.3	882	1	DPOL_SULSO		P26581 sulfobolus
30	616.5	15.3	1105	1	DPOL_MOUSE		P52431 mus musculus
31	15.3	1103	1	DPOL_MESOM		P97783 mesocricetus auratus	
32	614.5	15.3	1094	1	DPOL_PLAKF		P30315 plasmidium
33	611	15.2	1107	1	DPOL_HUMAN		P28340 homo sapiens

Scoring table: BLOSUM62 Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0% Maximum Match 100% Listing first 45 summaries

Database : SwissProt_41.4*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result ID	DPOL_THEST	STANDARD;	PRT; 1829 AA.
AC	Q33845;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	DNA polymerase (EC 2.7.7.7).		
GN	Thermococcus sp. (strain TY).		
OS	Archaeas; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.		
OC			
OX	NCBI TaxID=11063;		
RN	[1]		
RP	SEQUENCE FROM N.A. MEDLINE=98094267; PubMed=9434178;		
RA	Niehaus F., Frey B., Antzianikian G.; "Cloning and characterisation of a thermostable alpha-DNA polymerase from the hyperthermophilic archaeon Thermococcus sp. TY.";		
RT	Gene 204:152-158(1997).		
CC	- - CATALYTIC ACTIVITY: N deoxyribonucleoside triphosphate = N diphosphate + (DNA)N.		
CC	- - PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES (INTEINS) FOLLOWED BY PEPTIDE LIGATION.		
CC	- - SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL; Y1030; CPA73475.1; -.		
DR	HSS659; 1TGO.		
DR	InterPro; IPR0006172; DNA_Pol_B.		
DR	InterPro; IPR00134; DNA_Pol_B_dom.		
DR	InterPro; IPR005133; DNA_Pol_B_exo.		
DR	InterPro; IPR003586; Hedgehog_Ein.		
DR	InterPro; IPR003587; Hedgehog_hin.		
DR	InterPro; IPR006141; Interin.		
DR	InterPro; IPR006142; Interin.		
DR	InterPro; IPR004042; Interin_endonuc.		
DR	InterPro; IPR005178; Pol12.		
DR	Pfam; PF00136; DNA_Pol_B.		
DR	Pfam; PF03104; DNA_Pol_B_exo.		
DR	PRINTS; PRO0379; INTEIN.		
DR	SMART; SM00305; HintC.		
DR	SMART; SM00306; HintN.		
DR	TIGRFAMS; TIGR01443; Interin_Cterm.		3.
DR	TIGRFAMS; TIGR01445; Interin_Nterm.		3.
DR	TIGRFAMS; TIGR00592; Pol12.		2.
DR	PROSITE; PS00116; DNA_Polymerase_B; FALSE_NEG.		

ALIGNMENTS

DR	PROSITE; PS50818; INTEIN_N_TER; 3.	Db	VSPDTLREGCKNYDAPIVGKFFCKDFFGFFPSTLGEJITMRQEIKKMKATIDPIEKK 840
DR	PROSITE; PS50819; INTEIN_ENDONUCLEASE; 2.	Qy	481 MLDYRORAVLHA-----493
KW	Transferase; DNA-directed DNA Polymerase; DNA replication;	Db	841 MLDYRORAVLHS-----493
KW	DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;	Qy	494 -----493
KW	Protein splicing.	Db	901 VDNIAFSLNKESKSBKVKALLRHKYKGEAYEVELNSGRKIHITRGSLSFTIRNGKI 960
FT	CHAIN 1 409 DNA POLYMERASE, 1ST PART (POTENTIAL).	Qy	494 -----493
FT	CHAIN 410 769 INTEIN I.	Db	961 KBIWGEEVVKVGDLLIVPKKVLINEKEAVINIPELISKLPDDETDADVWMTPKGEKKNFFK 1020
FT	CHAIN 770 855 DNA POLYMERASE, 2ND PART (POTENTIAL).	Qy	494 -----493
FT	CHAIN 856 1392 INTEIN II.	Db	961 KBIWGEEVVKVGDLLIVPKKVLINEKEAVINIPELISKLPDDETDADVWMTPKGEKKNFFK 1020
FT	CHAIN 1393 1441 DNA POLYMERASE, 3RD PART (POTENTIAL).	Qy	494 -----493
FT	CHAIN 1442 1598 INTEIN III.	Db	961 KBIWGEEVVKVGDLLIVPKKVLINEKEAVINIPELISKLPDDETDADVWMTPKGEKKNFFK 1020
FT	CHAIN 1599 1829 DNA POLYMERASE, 4TH PART (POTENTIAL).	Qy	494 -----493
SQ	SEQUENCE 1829 AA; 211875 MW; A113ABBC57BB9CB3 CRC64;	Qy	494 -----493
Query Match Score 3446.5; DB 1; Length 1829;		Db	1021 GMRTJRKWIFGREGSKEIRTPNRYLFILEEGFVKLPLRGVEVTDWEGLKEYRQLYBKLVK 1080
Best Local Similarity 85.6%; Pred. No. 1.2e-179; Gaps 3;		Qy	494 -----493
Matches 771; Conservative 0; Mismatches 3; Indels 1055; Gaps 3;		Db	1081 NLRYNGNKRKREVLVRFNDIKDSVSCPKELEWKIGTXKGFRXXKICLKVDBDFGKFLGYY 1140
Qy	1 MIFDTDYITKDGKPIIIRIPIKKENGEPEKTEIDLPHPFOPYIYALLKODSADEIKAKGERHG 60	Db	1141 VSEGYYGAAQKNKTGCMSSVTKLYNNENPNVLDKDMKNTAEEKFFGKVVGKNCVTDIPKRMAYL 1200
Db	1 MILDQDITKDGKPIIIRIPIKKENGEPEKTEIDLPHPFOPYIYALLKODSADEIKAKGERHG 60	Qy	494 -----493
Qy	61 KIVRVAIVAKVYKVKKKFGLRDYEWVKLJFEHPDVPALRKIICREHPDVIYDIPFARRY 120	Db	1201 LAKSLCGYTAENKRIPIIIFDSSEPYRWAFLRAYFVGDGDIHPSKRLSLSTKSELLANQL 1260
Db	61 KIVRVAIVAKVYKVKKKFGLRDYEWVKLJFEHPDVPALRKIICREHPDVIYDIPFARRY 120	Qy	494 -----493
Qy	121 LIDKSLIPMGEDEELKUMAFDIEPYTHEGDSFGKGBIIMSYADBEARVITWNIDLPY 180	Db	1261 VFLINSLCVSSIKIGFDGDSGVYRVYINEDLPFLQTQSQRQNTTYPNLIPEKVELEIFGRKFQ 1320
Db	121 LIDKSLIPMGEDEELKUMAFDIEPYTHEGDSFGKGBIIMSYADBEARVITWNIDLPY 180	Qy	494 -----493
Qy	181 VDVVSNEREMIICPQVQVREXPDPVLTITYGDNFDLPLYLKRAEKLGVTLLGRDKHEPE 240	Db	1321 KNITFEKPKELADSQKDLRKYKLLDPLLNGDIVLDRVNTKRETYEGYYDLSVEDNEN 1380
Db	181 VDVVSNEREMIICPQVQVREXPDPVLTITYGDNFDLPLYLKRAEKLGVTLLGRDKHEPE 240	Qy	494 -----493
Qy	241 PKHMRGDSFAVEIKGHLFDFPVTVRTINPTLYTEAVFALGKTKSKLGAEFIAAI 300	Db	1381 FLYFGFLIYAHNSYGYGNGYKARWYSKECAESVTAWGRHYTEMTIKEBEKFGKLYA 1440
Db	241 PKHMRGDSFAVEIKGHRHDFLFPVTVRTINPTLYTEAVFALGKTKSKLGAEFIAAI 300	Qy	543 D-----543
Qy	301 WETBEESMKKLAQYSMEDARATYELGKEFFPMEAELAKLIGOSWMDVRSSTGNLYEWYLL 360	Db	1441 DSVTGDTEILVNRNGRTRIEFPVTEKLFERVDYRIGKEYCILEDVEALTLDRKGKLIWKKV 1500
Db	301 WETBEESMKKLAQYSMEDARATYELGKEFFPMEAELAKLIGOSWMDVRSSTGNLYEWYLL 360	Qy	544 -----543
Qy	361 RYAYERNELAQNPKDEEEYRRRLRTTLLGGYKYPEPEGLWNTITYDFRC-----410	Db	1501 PYMRHRRAKCKCYRITWNSWYDVTEDHSLLVAEDGKKEARPMETEGKSJLJATKDDSL 1560
Db	361 RYAYERNELAQNPKDEEEYRRRLRTTLLGGYKYPEPEGLWNTITYDFRC-----410	Qy	544 -----543
Qy	411 -----410	Db	1561 VEYKPHAEIEISYNGYYDIEVEETAKYQVPELEXVHETOITKDLSEYKAIAGPHVAI 685
Db	421 GKGIVNISDVKEGDIYLGIGWORVKWVKHTEGKLJININGLKC7PNHKPVV7ENDRQ 480	Qy	566 FLKINTSKPLGLLIEYLGFYLRGFVAKKRATVDEGTRITRGLEVVRDWSLAKET 625
Qy	411 -----410	Db	1621 FLKINTSKPLGLLIEYLGFYLRGFVAKKRATVDEGTRITRGLEVVRDWSLAKET 1680
Db	481 TRIRDSTAKSFLSGCKVKSKIIITKLFFKEXIAEFEKPKNPKSEEEILKGELSGIIAEGTLRK 540	Qy	626 QAKVTEAIIKEDSYKATEVETIVDVEETAKYQVPELEXVHETOITKDLSEYKAIAGPHVAI 685
Qy	411 -----410	Db	1681 QAKVTEAIIKEDSYKATEVETIVDVEETAKYQVPELEXVHETOITKDLSEYKAIAGPHVAI 1740
Db	541 DIEYFDSRGKKRISQYRVEITGENEKEELLERILYIFDKLFGIRPSVKRKGDNTALKI 600	Qy	686 AKRLAAKGKIKVRPSTIISIVLPGSGKISDRVLLSEYDPKHKYDPPYYTENQVLPVYL 745
Qy	411 -----410	Db	1741 AKRLAAKGKIKVRPSTIISIVLPGSGKISDRVLLSEYDPKHKYDPPYYTENQVLPVYL 1800
Db	661 KLLDSLGIPSYRSYKXINGKELTHILEITGRDGLLIFQTLVGFISSEKNEALEKAE 720	Qy	746 RILFAFGYKEKDIKQSSKQVGLDNLK 774
Qy	411 -----410	Db	1801 RILFAFGYKEKDIKQSSKQVGLDNLK 1829
Db	721 VREMRNLKNSFYNLSTFEVSSEYVGEVYDITLEGGNPYFANGILTHNSLPSI 780	Qy	RESULT 2
Qy	421 VSPDTLREGCKNYDAPIVGKFFCKDFFGFFPSTLGEJITMRQEIKKMKATIDPIEKK 480	DPOL_THELI	

ID	DPOL_THELI	STANDARD;	PRT;	1702 AA.	
AC	P30317;				
DT	01-APR-1993 (Rel. 25, Created)				
DT	01-APR-1993 (Rel. 41, Last annotation update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	DNA Polymerase (EC 2.7.7.7) (Vent DNA Polymerase) [Contains: Endonuclease PI-TIII (EC 3.1.-.-) (Tli pol-1 intein) (IVPS2); DE Endonuclease PI-TIII (EC 3.1.-.-) (Tli pol-2 intein) (IVPS1)].				
DE	POI, Endonuclease; Endonuclease PI-TIII (EC 3.1.-.-) (Tli pol-1 intein) (IVPS1);				
GN	Thermococcus litoralis.				
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococaceae; Thermoococcus.				
OC	NCBI_TAXID=2265;				
RN	[1]				
RP	SEQUENCE N.A., AND PARTIAL SEQUENCE.				
RX	MEDLINE=92302285; PubMed=1608969;				
RA	Perler F.B., Comb D.G., Jack W.E., Moran L.S., Qiang B., Carlow C.K.S., Jannasch H., Nwankwo D.O., Hempstead S.K., "Intervening sequences in an Archaea DNA Polymerase Gene."; Proc. Natl. Acad. Sci. U.S.A. 89:5577-5581(1992).				
RN	[2]				
RP	PROTEIN SPlicing.				
RX	Medline=3117083; PubMed=1415179;				
RA	Hodges R.A., Perler F.B., Noren C.J., Jack W.E.; "Protein splicing removes intervening sequences in an archaea DNA polymerase"; Nucleic Acids Res. 20:6153-6157(1992).				
CC	-!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.				
CC	-!- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE SITE-SPECIFIC RECOMBINATION INITIATED BY ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE INTEIN.				
CC	-!- BIOTECHNOLOGY: Used in the PCR method because of its high thermostability and low error rate. Sold by New England Biolabs.				
CC	-!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY + (DNA) (N).				
CC	-!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION (INTEINS), FOLLOWED BY PEPTIDE LIGATION.				
CC	-!- BIOTECHNOLOGY: Used in the PCR method because of its high thermostability and low error rate. Sold by New England Biolabs.				
CC	-!- SIMILARITY: IN THE INTEIN SECTION? BELONGS TO THE HOMING ENDONUCLEASE FAMILY.				
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce or send an email to license@isb-sib.ch).				
CC	EMBL; M74198; AAA72101.1;				
DR	PIR; S42459; S42459.				
DR	HSSP; P56689; 1TGO.				
DR	REBASE; 2613; PI-TIII.				
DR	InterPro; IPR06172; DNA_Pol_B.				
DR	InterPro; IPR06134; DNA_Pol_B_dom.				
DR	InterPro; IPR06133; DNA_Pol_B_exo.				
DR	InterPro; IPR03586; Hedgehog_HintC.				
DR	InterPro; IPR03587; Hedgehog_hintN.				
DR	InterPro; IPR06141; Intein.				
DR	InterPro; IPR06142; Intein.				
DR	InterPro; IPR04542; Intein_endonuc.				
DR	InterPro; IPR04578; Pol2.				
DR	Pfam; PF00136; DNA_Pol_B.				
DR	Pfam; PF03104; DNA_Pol_B_exo.				
DR	PRINTS; PR00319; INTEIN.				
DR	SMART; SM00305; HintC.				
DR	SMART; SM00306; HintN.				
DR	SVRNGNKRFLVLMFNEIKDFISYPQKLEELWKGITLNGPRNCILKVDDEFQKLGY				
DR	SMART; SM00486; POLBC.	1.			
DR	TIGRFAMS; TIGR01443; intein_Cterm.	2.			
DR	TIGRFAMS; TIGR01445; intein_Nterm.	2.			
DR	TIGRFAMS; TIGR0592; PC02; 2.				
DR	PROSITE; PS00116; DNA POLYMERASE_B; FALSE_NEG.				
DR	PROSITE; PS50818; INTEIN_CTER; 2.				
DR	PROSITE; PS50819; INTEIN_ENDONUCLEASE; 2.				
DR	PROSITE; PS50817; INTEIN_NTER; 2.				
KW	Transferase; DNA-directed DNA polymerase; DNA replication;				
KW	DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;				
KW	Protein splicing; Intron homing.				
FT	CHAIN 1	494			DNA POLYMERASE, 1ST PART.
FT	CHAIN 1	1032			ENDONUCLEASE_P1-TLI_11.
FT	CHAIN 1	1081			DNA POLYMERASE, 2ND PART.
FT	CHAIN 1	1471			ENDONUCLEASE_P1-TLI_1.
FT	CHAIN 1	1702			DNA POLYMERASE, 3RD PART.
SQ	SEQUENCE 1702 AA;	197293 MW;	216598C75F53B0	CRC4;	
Query	1 MIFDDYDITKDGKP1IRIFKENGKFIELDPHFQPYVALLQDSDAIDBIKAIKGHRG	60			Score 3356; DB 1; Length 1702;
Query	1 MILDTDYITKDGKP1IRIFKENGKFIELDPHFQPYVALLQDSDAIDBIKAIKGHRG	60			Best Local Similarity 42.7%; Pred. No. 8.7e-175;
Matches	726;	Conservative 33;	Mismatches 15;	Indels 928;	Gaps 2;
Query	1 KIVRTVDAVKYKKEPLGRDVBWKL1FEHQDVBALGKIREPAVIDYBYDIPFARRY	120			
Db	1 KIVRTVDAVKYKKEPLGRDVBWKL1FEHQDVBALGKIREPAVIDYBYDIPFARRY	120			
Query	1 LIDKGKLMPEGEDEELKLMFAEDIETYHEGDEFKGKETIMSYADEEARVTKWIDLY	180			
Db	1 LIDKGKLMPEGEDEELKLMFAEDIETYHEGDEFKGKETIMSYADEEARVTKWIDLY	180			
Query	1 LIDKGKLMPEGEDEELKLMFAEDIETYHEGDEFKGKETIMSYADEEARVTKWIDLY	180			
Db	1 LIDKGKLMPEGEDEELKLMFAEDIETYHEGDEFKGKETIMSYADEEARVTKWIDLY	180			
Query	181 YDVSISMERENIKRFQIVREKDPOVLYITQDNLPLYIKRAEKGIGTLLGRDEKEHE	240			
Db	181 YDVSISMERENIKRFQIVREKDPOVLYITQDNLPLYIKRAEKGIGTLLGRDEKEHE	240			
Query	241 PKIHRNGDSFAVEIKGRIHFDLFPVTRTINLPYTLPEAVYAVLGIKTKSKLGAEI	300			
Db	241 PKIHRNGDSFAVEIKGRIHFDLFPVTRTINLPYTLPEAVYAVLGIKTKSKLGAEI	300			
Query	301 WETEESMKKLAQYSMEDARATYELGKEPPMBAEAKLIGQSWDVSRSSTGMLVWILL	360			
Db	301 WETEESMKKLAQYSMEDARATYELGKEPPMBAEAKLIGQSWDVSRSSTGMLVWILL	360			
Query	361 RVAYERNELLAPNKPDEBEYRRRLTYLGYYKEPGLWNTYDRCLYPSITVTHN	420			
Db	361 RVAYERNELLAPNKPDEBEYRRRLTYLGYYKEPGLWNTYDRCLYPSITVTHN	420			
Query	421 VSPDTLREGCKNYDAPIVGYKFCDFGGFIPSILGELITWROEIKKMKATIDPIEKK	480			
Db	421 VSPDTLREGCKNYDAPIVGYKFCDFGGFIPSILGELITWROEIKKMKATIDPIEKK	480			
Query	481 MLDYRORAYKLHA-----	493			
Db	481 MLDYRORAYKLHA-----	493			
Query	541 VNNLIAFSNKKIKKSEVKVKALTRHKYKGAVYIQLSSGRKINTAGHSFLTVRNGE	600			
Db	541 VNNLIAFSNKKIKKSEVKVKALTRHKYKGAVYIQLSSGRKINTAGHSFLTVRNGE	600			
Query	494 -----	493			
Db	601 KEVSG3DGKIKEGDLIVYAPKKKLNEKGVSINTPELISDLSSEETADIVMTISAKGRKNPFK	660			
Query	494 -----	493			
DR	721 SVRNGNKRFLVLMFNEIKDFISYPQKLEELWKGITLNGPRNCILKVDDEFQKLGY	780			

Qy	494	-----	493	-----	DE DNA polymerase (EC 2.7.7.7) (TO POL).
Db	781	VSEGYAAQQRNKTGGISYSVKNLYNEDDNVLSMRNVALEKFFGKVRDNCVTSISKMAYL	840	GN OS Thermococcus gorgonarius.	
Qy	494	-----	493	-----	OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
Qy	494	-----	493	-----	NCBI_TaxID=71997;
Db	B 41	VMKCLCGALAENKRIPPSVILTSPEPVWSFLEAYFTGDTHPSKRFLSTKSELLANQL	900	RN [1] X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).	
Qy	494	-----	493	-----	RP MEDLINE-99192230; PubMed=10097083;
Db	901	VFLINSLGISSYKIGFDGKVRYVINYNDLQFFQTSREKRNTTYSNLPIKEIIRDVFGKEFQ	960	RX RA Hopfner K.-P., Eichinger A., Engl R.A., Lueb F., Ankenbauer W.,	
Qy	494	-----	493	-----	RA Huber R., Angerer B.; "Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius".
Db	961	KMTFKPKFKEVYDGSKLNRKAKLLEFFINGDIVLDRKSYKEKDYGYYDLSVDENEN	1020	RT Proc. Natl. Acad. Sci. U.S.A. 96:3600-3605 (1999).	
Qy	494	-----	493	-----	RL - - FUNCTION IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE EXHIBITS 3', TO 5', EXONUCLEASE ACTIVITY.
Db	1021	FLVGFGLLYAHNSYYGGMGYPKARYSKCAESVTA	1080	CC CC - - CATALYTIC ACTIVITY: N deoxyribonucleoside triphosphate = N diphosphate + [DNA] (N).	
Qy	543	D-----	543	-----	CC - - SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
Db	1081	DVSYGSEIILRQNGKBRFKVTKDPLKXVDSIGERKEYCILEGV	1140	DR DR PDB: 1TGO; 22-MAR-99.	
Qy	544	-----	543	-----	DR DR InterPro: IPRO06134; DNA_Pol_B_dom.
Db	1141	PTYMRHANKRMFRILWNTNSWYIDVTDHSIGLYLNTSKTAKIGERLKEV	1200	DR DR InterPro: IPRO04578; PolB.	
Qy	544	-----	543	-----	DR DR Pfam: PF00136; DNA_Pol_B_1.
Db	1201	AVKSLICPNAPLDENTTKTSETIAVKFVELVGLIVGDDNWGGDSRWA	1260	DR DR Pfam: PF03104; DNA_Pol_B_exo; 1.	
Qy	544	-----	543	-----	DR DR PRANTS; PRO00106; DNAPOLB.
Db	1261	IKQXLEPLPKTYGVISNYPKNEKGDFNLLAKSLVKMFKDFEXRKRTPEFMYLPV	1320	DR DR SM00466; POLB; 1.	
Qy	544	-----	543	-----	DR DR TIGRFAMS; TIGR00512; PolB; 1.
Db	1321	TYIEAFRLGLFSAADGTVTIRKGVPEIRLTNTADAF-REVRLWIVGINSIFAETTPNR	1380	DR DR PROSITE; PS00116; DNA POLYMERASE_B; 1.	
Qy	544	-----	543	-----	DR DR TRANSFERASE; DNA-directed DNA Polymerase; DNA replication;
Db	1381	YNGVSTGTYSKHLRINKWKRFAERIGFLIERQKRLLEHLSARV	1440	KW KW DNA-binding; Hydrolase; Nuclease; Exonuclease; Multifunctional enzyme;	
Qy	544	-----	543	-----	FT FT 3D-STRUCTURE
Db	1441	KVEEPIPYEGVYDIEVEETHFANNLVNNDGFYATIPEKPKLKKRKAKEFLYINS	1500	FT FT DISULFID 428 442	
Qy	573	KLPOLLLELEYEGFLRGFFVAKRYAVIDEDEGRITGGLEVRRDWSSEIAKEOAKVLEA	1532	FT FT DISULFID 506 509	
Db	1501	KLPOLLLELEYEGFLRGFFVAKRYAVIDEDEGRITGGLEVRRDWSSEIAKEOAKVLEA	1560	FT FT STRAND 2 10	
Qy	633	ILKEDSTVKAEEVKDQVVEIAKCYVCPLEKVIHETQTKDSEYKAIGPHAJAKPLAAK	692	FT FT TURN 11 12	
Db	1561	ILKEGSVEAKEVYDVRDVVEKIAKYRVPLEKLVHETQTRDLDYKAGPHAJAKLAA	1620	FT FT STRAND 13 22	
Qy	693	GIKYRPGTISIISYVLRGSKGKISDRVILLSEYDPKGKHYDPDYYIENQYLPAVTRILEAFG	752	FT FT TURN 23 24	
Db	1621	GIKYRPGTISIISYVKGSKRISDRVILLSEYDPKGKHYDPDYYIENQYLPAVTRILEAFG	1680	FT FT STRAND 25 31	
Qy	753	YRKEDLKYQSSKQVGLDAWLLK 774		FT FT STRAND 37 42	
Db	1681	YRKEDLRYQSSKQVGLDAWLLK 1702		FT FT STRAND 47 47	
				FT FT HELIX 48 51	
				FT FT STRAND 52 53	
				FT FT STRAND 55 58	
				FT FT TURN 59 60	
				FT FT STRAND 61 64	
				FT FT TURN 67 75	
				FT FT HELIX 76 77	
				FT FT STRAND 78 86	
				FT FT TURN 90 91	
				FT FT HELIX 92 102	
				FT FT STRAND 104 105	
				FT FT TURN 106 110	
				FT FT HELIX 116 123	
				FT FT STRAND 124 125	
				FT FT STRAND 137 144	
				FT FT TURN 152 153	
				FT FT STRAND 205 208	
				FT FT HELIX 211 213	
				FT FT TURN 214 214	
				FT FT HELIX 215 225	
				FT FT TURN 226 227	
				FT FT STRAND 234 243	
				FT FT STRAND 240 240	
				FT FT STRAND 248 251	
				FT FT TURN 253 253	
				RESULT 3	
				DPOL_TheGO	
				ID_DPOL_TheGO	
				STANDARD	
				PRT; 773 AA.	
				AC P56619;	
				DT 15-JUL-1999 (Rel: 38, Created)	
				DT 15-JUL-1999 (Rel: 38, Last sequence update)	
				DT 28-FEB-2003 (Rel: 41, Last annotation update)	

		Matches	605;	Conservative	83;	Mismatches	82;	Indels	3;	Gaps	2;
FT	STRAND	256	259								
FT	HELIX	267									
FT	HELIX	275	283								
FT	HELIX	292	301								
FT	TURN	303	304								
FT	HELIX	305	337								
FT	TURN	338									
FT	HELIX	341	345								
FT	TURN	346	346								
FT	HELIX	349									
FT	TURN	364	365								
FT	STRAND	366	366								
FT	HELIX	374	379								
FT	STRAND	389	390								
FT	STRAND	396	405								
FT	HELIX	408	415								
FT	TURN	416	417								
FT	STRAND	420	422								
FT	STRAND	423	423								
FT	TURN	424									
FT	TURN	426									
FT	STRAND	431	433								
FT	TURN	435	437								
FT	STRAND	440	442								
FT	HELIX	448									
FT	HELIX	473	490								
FT	TURN	491	492								
FT	HELIX	493	498								
FT	TURN	500	501								
FT	TURN	503	504								
FT	HELIX	507	530								
FT	TURN	531	532								
FT	STRAND	535	540								
FT	STRAND	543	547								
FT	HELIX	553	550								
FT	TURN	569	570								
FT	STRAND	573	574								
FT	TURN	577									
FT	TURN	591	592								
FT	STRAND	593	597								
FT	TURN	599	600								
FT	STRAND	603	606								
FT	STRAND	615	615								
FT	HELIX	617	631								
FT	TURN	632	633								
FT	HELIX	636	651								
FT	TURN	652	653								
FT	HELIX	657	659								
FT	STRAND	662									
FT	HELIX	670	672								
FT	HELIX	679	689								
FT	TURN	690	691								
FT	TURN	695	696								
FT	STRAND	698	704								
FT	TURN	710	711								
FT	STRAND	714	716								
FT	HELIX	717	719								
FT	TURN	722	724								
FT	STRAND	727	727								
FT	HELIX	729	735								
FT	TURN	736	737								
FT	HELIX	738	740								
FT	TURN	741	741								
FT	HELIX	742	746								
FT	TURN	747	749								
FT	HELIX	752	754								
FT	TURN	766	767								
FT	TURN	771	773								
SEQ	SEQUENCE	773	AA:	89812	MW:	F67AFF04E875FBE44	CRC64;				
Query Match				80.8%		Score 3253.5;	DB 1;	Length 773;			
Best Local Similarity				78.3%		Pred. No. 1.2e-169;					

RESULT 4
 DPOL_THESS9
 ID _DPOL_THESS9
 STANDARD; PRT; 775 AA.
 AC Q56336;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-SEP-2003 (Rel. 37, Last sequence update)
 DE DEDDKitTRGLEIVRDWSEJAKETQARVLEIATKRGDVEAIVRIVKEVTEKLSKYEVP 657
 661 EKLVTHQIKDLSYKAKTGAQPHVAKLAKAAGKTKVRPOTIISVILRSQGKISDRVILL 720
 DB 658 EKLVYEQITRDLKHYKATSPHYAVAKRUAARGKIRPTEVISTVILRSQGKISDRVILL
 601 DEEGRITTRGLEIVRDWSEJAKETQAKVLEIATKEDSVKAVIKDQVVEETIAKQYVPL 660
 DB 598 DEEDDKitTRGLEIVRDWSEJAKETQARVLEIATKRGDVEAIVRIVKEVTEKLSKYEVP 657
 721 SEYDDKKHCYDPDYIENQVLPAYRILPAGFTRKEDLYKQSSKQVGDWK 773
 QY 718 DEFDAKHYDAEYTIENQVLPAYRILPAGFTRKEDLYKQSSKQVGDWK 770
 OS Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococceae;
 OC Thermococcus
 NCBI_TaxID=101799;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96224274; PubMed=8643567;
 RA Southworth M.W., Kong H., Rucera R.B., Ware J., Jannasch H.W.,
 RA Perier F.B.,

"Cloning of thermostable DNA Polymerases from hyperthermophilic marine Archaea with emphasis on Thermococcus sp. 9ON-7 and mutations affecting 3', 5' exonuclease activity."

RT Proc. Natl. Acad. Sci. U.S.A. 93:5281-5285 (1996).

RL + [DNA] (N).

CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to licensee@isb-sib.ch).

DR BMBL; U47108; AAAB88169_1; -.

DR PIR; S67920; S67920; -.

DR PDB; 1QHT; 04-TUN-00; -.

DR InterPro; IPR006172; DNA_Pol_B.

DR InterPro; IPR006134; DNA_Pol_B_exo.

DR InterPro; IPR006133; DNA_Pol_B_exo.

DR InterPro; IPR004578; Pol2.

DR Pfam; PF00136; DNA_Pol_B_1.

DR Pfam; PF003104; DNA_Pol_B_1.

DR PRINTS; PR00106; DNAPoB.

DR TIGRFAMS; TIGR0592; PolB_1.

DR PROSITE; PS00016; DNA POLYMERASE_B_1.

DR DNA Transferase; DNA-directed DNA Polymerase; DNA replication;

DR DNA-Binding; 3D-structure.

DR SEQUENCE 775 AA: 89781_MW: 71525D63C09C9D9 CRC64;

DR Query Match 79.6%; Score 3204.5; DB: 1; Length 775; Best Local Similarity 76.6%; Pred. No. 5.4e-16; Indels 3; Gaps 2; Matches 592; Conservative 89; Mismatches 89; -.

DR 1 MIFDTDTIKTDKSPITIPIRKENGEEFKIILDPHFQPYTYALLKODSAIDBIKALKERHG 60

DR 1 MILDTDYTENGKPVIRVKPVGKENGEEFKIYEYDRTFEPYFYALKDSSAIEVKYTAKHG 60

DR 61 KIVRVPDAVKVKKFLGRDVEYKFLIFEPHQCPVPAIPLRKTRHPAVTDYEYDIFPARRY 120

DR 61 TVVKVKRKAERVKQKKFLGRPPIEWKLYFVNHQDVPVARDRRAHPAVDYEYDIFPARRY 120

DR 121 LIDKGILPMEGDEBKLMAFDIETPYHEGDFEGKGEITIMSYADEBEARYTWNIDLPY 180

DR 121 LIDKGILPMEGDEBKLMAFDIETPYHEGDFEGKGEITIMSYADEBEARYTWNIDLPY 180

DR 181 VDVVSNSREMIKRPVQIYREKDPDVITYLIGKRAEKUGGTVLGLRDKHEPHE 240

DR 181 VDVVSREMIKRPVQIYREKDPDVITYLIGKRAEKUGGTVLGLRDKHEPHE 240

DR 241 PKTHRMGGSFAVEIKGTHFOLPPVTRTNPETYIYEAVLGKTKSGLGAEIIAI 300

DR 239 PKLQRMGDRFAVKGTHFOLPPVTRTNPETYIYEAVLGKTKSGLGAEIIAI 298

DR 301 WETEESMKKLAQSMEDARATELGKRFPPMEEALKLIGQSVWDVRSSTGNYEWYLL 360

DR 299 WESGEGLERVARSMEDAKVTLGRFPEGLWNTYDFFAYLKRCEELGKFTLGRDGS--E 358

DR 361 RAVERNELAQNPKDDEEYRRRLRTTIGGGYKPEGLWNTYDFFAYLKRCEELGKFTLGRDGS--E 420

DR 359 RKA YKRNELLAPNPKDDERELARR-RGGAGGGYKPERGLWDNIVLDRSLYSPSITIEN 417

DR 421 VSPDTLERIGCKKNYDVAIIVGYKFCKDPGFTISLGLPITYIYEAVLGKTKSGLGAEIIAI 480

DR 418 VSPDTLNRGCKKEYDVAPIVGHKFCDPFFGFISSLGLDLEERQKFRRMKATVDLEKK 477

DR 481 MLDYRQRAVKLHANSYGMGKPARWSKECAESVTAWGREYIEMTIKEIIEKFGFKV 540

DR 478 LLDYRQRAIKLANSFYGTYGAKARYWCKECAESVTAWGREYIEMTIKEIIEKFGFKV 537

Qy 541 YADTDGFYATIPIGKPKETIKKAKFLKLYINSKLPGLLEBEFGYLRGFFVAKKRYAVI 600

Db 538 YADTDGLHATIPGDAETVKKAKFLKLYINSKLPGLLEBEFGYLRGFFVAKKRYAVI 597

Qy 601 DEEGKTTTRGLEYVWRDWSIAKETQAKYLEAILEDSEVAKBVIKDVYETAKYQVPL 660

Db 598 DEEGKTTTRGLEYVWRDWSIAKETQAKYLEAILEDSEVAKBVIKDVYETAKYQVPL 657

Qy 661 EKLVTHEQITMDLSYKAIAPHVATAKRLAAKGIKVPGMIIISVLRGSGKISDRVILL 720

Db 658 EKLVTHEQITMDLSYKAIAPHVATAKRLAAKGIKVPGMIIISVLRGSGKISDRVILL 717

Qy 721 SEYDPKKHKKYDPDYYTIENQVLPAVYRILAFGTYREDLQYSSKOGVGDRAIPA 773

Db 718 DEFDPKTHRDAEYXTIENQVLPAVYRILAFGTYREDLQYSSKOGVGDRAIPA 770

CC RESULT 5

CC DPOL_PYRAB STANDARD; PRT: 771 AA.

CC ID DPOL_PYRAB

CC AC P77916; P77932; 37, Created

CC DT 15-DEC-1998 (Rel. 37, Last sequence update)

CC DT 30-MAY-2000 (Rel. 39, Last sequence update)

CC DT 15-SEP-2003 (Rel. 42, Last annotation update)

CC DB DNA_polymerase_1 (EC 2.7.7.7) (Pub Polymerase).

CC GN POL OR POL_OR PYRAB17200 OR PAB128.

CC OS Pyrococcus abyssi

CC Archaea; Euryarchaeota; Thermococci; Thermococccaceae;

CC Pyrococcus

CC NCBI_TAXID=29292;

CC RN [1]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=ST 855;

CC RA Carbon M., Querellou J., Barbier G., Dietrich J.-t. Forterre P.;

CC RA Carbon M., Querellou J., Barbier G., Dietrich J.-t. Forterre P.;

CC RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC RN [2]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=GE5 / Orsay;

CC RX Published=12622808;

CC RA Cohen G.N., Barbe V., Filament D., Galperin M., Heilig R., Leecompte O., Poeh O., Prieur D., Querellou J., Ripp R., Thierry J.-C., Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;

CC RA Forterre P., Dietrich J.-t. Forterre P.;

CC RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC RL [3]

CC RN [4]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=GE5 / Orsay;

CC RX Published=12622808;

CC RA Cohen G.N., Barbe V., Filament D., Galperin M., Heilig R., Leecompte O., Poeh O., Prieur D., Querellou J., Ripp R., Thierry J.-C., Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;

CC RA Forterre P., Dietrich J.-t. Forterre P.;

CC RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC RL [5]

CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Z34174; CAA00888_1; -.

CC EMBL: Z51473; CAA00887_1; -.

CC DR PIRBL; C75023; C75023.

CC DR HSSP; P56689; 1TGO.

CC DR InterPro; IPR006172; DNA_pol_B.

CC DR InterPro; IPR006134; DNA_pol_B.

CC DR InterPro; IPR006133; DNA_pol_B_exo.

CC DR InterPro; IPR004578; Pol2.

CC DR PFM; PF00136; DNA_Pol_B_1.

CC DR DR

DR Pfam: PF03104; DNA pol B - exc; 1.

DR PRNTS; PRO016; DNAPOFB.

DR SMART; SM00488; POLBc; 1.

DR TIGRFAMS; TIGR00522; POL2; 1.

DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.

DR TRANSFERASE; DNA-directed DNA polymerase; 1.

DR KWD DNA-binding; Complete proteome.

FT VARIANT 532 532 R -> S (IN STRAIN GE23).

FT VARIANT 553 554 HE -> P (IN STRAIN GE23).

FT VARIANT 756 756 K -> R (IN STRAINS GE23 AND ST 855).

FT CONFLICT 263 263 V -> A (IN REF. 1).

FT CONFLICT 277 277 A -> T (IN REF. 1).

FT CONFLICT 281 281 A -> V (IN REF. 1).

FT CONFLICT 320 320 F -> S (IN REF. 1).

FT CONFLICT 339 339 Q -> H (IN REF. 1).

FT CONFLICT 359 359 R -> T (IN REF. 1).

FT CONFLICT 391 391 K -> N (IN REF. 1).

SEQUENCE 771 AA; 83496 NW; 110A87045A8A5522 CRC64;

Query Match 78.6%; Score 3164.5; Best Local Similarity 76.1%; Matches 388; Conservative 89; Mismatches 93; Indels 3; Gaps 2;

Db 1 MIFDPTDYYTDKGPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEIKAKGERHG 60

Db 1 MTDADYTDGKPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEVKRITAHRG 60

Qy 61 KIVTRVDAVKVKKFKLGRDVEWKLIFHPDQVPAFLRGKIKRHPAVDIDYDIPPAKRY 120

Db 61 KIVTRVTEVKVQCKFLPPIEWKVLHLPQDQVPAFKRKHHPAVDIDYDIPPAKRY 120

Qy 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Db 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Qy 181 VDVSNEERMIKEPKVQIYREKDPDVLYTNGNFDLDPYLKRAKGUTVLLGRDKHEHPE 240

Db 181 VEVSSSEPKMILKLAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Qy 241 PKHRMGDSFAVJIKRHFDFPVVERTINPFTYDLEAVYTAVLTKSKLIGAEIIAI 300

Db 239 PKNCRMGSLAVBKGRHFDLFPVIRRTINPFTYDLEAVYTAFLGKSKVETVAAEA 298

Qy 301 WEFBESMCKLAAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Db 299 WEFCKGLPRAVXYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 358

Qy 361 RVAYERNLAPNPKPDEEFYRRLRTTLYGGYKTEPERGLWENITYLDFRCLYPSLIVTHN 420

Db 359 RKAYERNLAPNPKPDERYERLIRESEGGYKTEPERGLWGIVSDFRSRSTPSLITTHN 418

Qy 421 VSPDTLREGCKYDVAFLVGYKFCCKGKPFGETPSLIGELIMMQEIKKKKATIDIEKK 480

Db 419 VSPDTLNFENCKYDVAQGFRCKDFPGFPPSLGNLLEERQKTKRMKESDFVEKK 478

Qy 481 MLDYRORAVKLAHANSYCCYGMGPKARYTKCKEASSTWAGRYHITEMKIEEKFGKFL 540

Db 479 LLDDYRQKIKLANSYCCYGMGPKARYTKCKEASSTWAGRYDILVRELESR-GFKTL 537

Qy 541 YADTDGFATIPEKPKTIIKAKFELKFKYKINSKLPGILEEYGFYLRGKIKRKYAVI 600

Db 548 YIPTDGLYATIPIGAKHEEIKERALKFVKEYINSKLPGILEEYGFYLRGKIKRKYAVI 597

Qy 601 DESGRITRGLYEVRRDSEIATKETQAKVBLAIKEDSVEKAVETYKDVVEIACQVPL 660

Db 598 DEEGKIVTRGLEYVRRDSEIATKETQAKVBLAIKGNVDEAVKIVKEVTEKLSKEIIP 657

Qy 661 EKLVTHEQITKOLSEYKAGKPHVIAAKRLAAGKIKYRGTISYVLRGSGKISDRVILL 720

Db 658 EKLVYEQITRPLSEYKAGPHVIAAKRLAAGKVKYRGPWVYIVLRGDOPISKRAIAI 717

Qy 721 SEYDPKKHCKYDQDYYTENQVTPAVLRLTEAFGSKRDKLKYQSSKQVGLDAWIK 773

DR 718 BFFDPKKHCKYDAEYVYIENQVLPAVERILRAFGTRKEDLYQKTKQVGIGAWLK 770

DR SMART; SM00488; POLBc; 1.

DR TIGRFAMS; TIGR00522; POL2; 1.

DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.

DR TRANSFERASE; DNA-directed DNA polymerase; 1.

DR KWD DNA-binding; Complete proteome.

FT VARIANT 532 532 R -> S (IN STRAIN GE23).

FT VARIANT 553 554 HE -> P (IN STRAIN GE23).

FT VARIANT 756 756 K -> R (IN STRAINS GE23 AND ST 855).

FT CONFLICT 263 263 V -> A (IN REF. 1).

FT CONFLICT 277 277 A -> T (IN REF. 1).

FT CONFLICT 281 281 A -> V (IN REF. 1).

FT CONFLICT 320 320 F -> S (IN REF. 1).

FT CONFLICT 339 339 Q -> H (IN REF. 1).

FT CONFLICT 359 359 R -> T (IN REF. 1).

FT CONFLICT 391 391 K -> N (IN REF. 1).

SEQUENCE 771 AA; 83496 NW; 110A87045A8A5522 CRC64;

Query Match 78.6%; Score 3164.5; Best Local Similarity 76.1%; Matches 388; Conservative 89; Mismatches 93; Indels 3; Gaps 2;

Db 1 MIFDPTDYYTDKGPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEIKAKGERHG 60

Db 1 MTDADYTDGKPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEVKRITAHRG 60

Qy 61 KIVTRVDAVKVKKFKLGRDVEWKLIFHPDQVPAFLRGKIKRHPAVDIDYDIPPAKRY 120

Db 61 KIVTRVTEVKVQCKFLPPIEWKVLHLPQDQVPAFKRKHHPAVDIDYDIPPAKRY 120

Qy 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Db 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Qy 181 VDVSNEERMIKEPKVQIYREKDPDVLYTNGNFDLDPYLKRAKGUTVLLGRDKHEHPE 240

Db 181 VEVSSSEPKMILKLAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Qy 241 PKHRMGDSFAVJIKRHFDFPVVERTINPFTYDLEAVYTAVLTKSKLIGAEIIAI 300

Db 239 PKNCRMGSLAVBKGRHFDLFPVIRRTINPFTYDLEAVYTAFLGKSKVETVAAEA 298

Qy 301 WEFBESMCKLAAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Db 299 WEFCKGLPRAVXYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 358

Qy 361 RVAYERNLAPNPKPDEEFYRRLRTTLYGGYKTEPERGLWENITYLDFRCLYPSLIVTHN 420

Db 359 RKAYERNLAPNPKPDERYERLIRESEGGYKTEPERGLWGIVSDFRSRSTPSLITTHN 418

Qy 421 VSPDTLREGCKYDVAFLVGYKFCCKGKPFGETPSLIGELIMMQEIKKKKATIDIEKK 480

Db 419 VSPDTLNFENCKYDVAQGFRCKDFPGFPPSLGNLLEERQKTKRMKESDFVEKK 478

Qy 481 MLDYRORAVKLAHANSYCCYGMGPKARYTKCKEASSTWAGRYHITEMKIEEKFGKFL 540

Db 479 LLDDYRQKIKLANSYCCYGMGPKARYTKCKEASSTWAGRYDILVRELESR-GFKTL 537

Qy 541 YADTDGFATIPEKPKTIIKAKFELKFKYKINSKLPGILEEYGFYLRGKIKRKYAVI 600

Db 548 YIPTDGLYATIPIGAKHEEIKERALKFVKEYINSKLPGILEEYGFYLRGKIKRKYAVI 597

Qy 601 DESGRITRGLYEVRRDSEIATKETQAKVBLAIKEDSVEKAVETYKDVVEIACQVPL 660

Db 598 DEEGKIVTRGLEYVRRDSEIATKETQAKVBLAIKGNVDEAVKIVKEVTEKLSKEIIP 657

Qy 661 EKLVTHEQITKOLSEYKAGKPHVIAAKRLAAGKIKYRGTISYVLRGSGKISDRVILL 720

Db 658 EKLVYEQITRPLSEYKAGPHVIAAKRLAAGKVKYRGPWVYIVLRGDOPISKRAIAI 717

Qy 721 SEYDPKKHCKYDQDYYTENQVTPAVLRLTEAFGSKRDKLKYQSSKQVGLDAWIK 773

DR SMART; SM00488; POLBc; 1.

DR TIGRFAMS; TIGR00522; POL2; 1.

DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.

DR TRANSFERASE; DNA-directed DNA polymerase; 1.

DR KWD DNA-binding; Complete proteome.

FT VARIANT 532 532 R -> S (IN STRAIN GE23).

FT VARIANT 553 554 HE -> P (IN STRAIN GE23).

FT VARIANT 756 756 K -> R (IN STRAINS GE23 AND ST 855).

FT CONFLICT 263 263 V -> A (IN REF. 1).

FT CONFLICT 277 277 A -> T (IN REF. 1).

FT CONFLICT 281 281 A -> V (IN REF. 1).

FT CONFLICT 320 320 F -> S (IN REF. 1).

FT CONFLICT 339 339 Q -> H (IN REF. 1).

FT CONFLICT 359 359 R -> T (IN REF. 1).

FT CONFLICT 391 391 K -> N (IN REF. 1).

SEQUENCE 771 AA; 83496 NW; 110A87045A8A5522 CRC64;

Query Match 78.6%; Score 3164.5; Best Local Similarity 76.1%; Matches 388; Conservative 89; Mismatches 93; Indels 3; Gaps 2;

Db 1 MIFDPTDYYTDKGPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEIKAKGERHG 60

Db 1 MTDADYTDGKPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEVKRITAHRG 60

Qy 61 KIVTRVDAVKVKKFKLGRDVEWKLIFHPDQVPAFLRGKIKRHPAVDIDYDIPPAKRY 120

Db 61 KIVTRVTEVKVQCKFLPPIEWKVLHLPQDQVPAFKRKHHPAVDIDYDIPPAKRY 120

Qy 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Db 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Qy 181 VDVSNEERMIKEPKVQIYREKDPDVLYTNGNFDLDPYLKRAKGUTVLLGRDKHEHPE 240

Db 181 VEVSSSEPKMILKLAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Qy 241 PKHRMGDSFAVJIKRHFDFPVVERTINPFTYDLEAVYTAVLTKSKLIGAEIIAI 300

Db 239 PKNCRMGSLAVBKGRHFDLFPVIRRTINPFTYDLEAVYTAFLGKSKVETVAAEA 298

Qy 301 WEFBESMCKLAAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Db 299 WEFCKGLPRAVXYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 358

Qy 361 RVAYERNLAPNPKPDEEFYRRLRTTLYGGYKTEPERGLWENITYLDFRCLYPSLIVTHN 420

Db 359 RKAYERNLAPNPKPDERYERLIRESEGGYKTEPERGLWGIVSDFRSRSTPSLITTHN 418

Qy 421 VSPDTLREGCKYDVAFLVGYKFCCKGKPFGETPSLIGELIMMQEIKKKKATIDIEKK 480

Db 419 VSPDTLNFENCKYDVAQGFRCKDFPGFPPSLGNLLEERQKTKRMKESDFVEKK 478

Qy 481 MLDYRORAVKLAHANSYCCYGMGPKARYTKCKEASSTWAGRYHITEMKIEEKFGKFL 540

Db 479 LLDDYRQKIKLANSYCCYGMGPKARYTKCKEASSTWAGRYDILVRELESR-GFKTL 537

Qy 541 YADTDGFATIPEKPKTIIKAKFELKFKYKINSKLPGILEEYGFYLRGKIKRKYAVI 600

Db 548 YIPTDGLYATIPIGAKHEEIKERALKFVKEYINSKLPGILEEYGFYLRGKIKRKYAVI 597

Qy 601 DESGRITRGLYEVRRDSEIATKETQAKVBLAIKEDSVEKAVETYKDVVEIACQVPL 660

Db 598 DEEGKIVTRGLEYVRRDSEIATKETQAKVBLAIKGNVDEAVKIVKEVTEKLSKEIIP 657

Qy 661 EKLVTHEQITKOLSEYKAGKPHVIAAKRLAAGKIKYRGTISYVLRGSGKISDRVILL 720

Db 658 EKLVYEQITRPLSEYKAGPHVIAAKRLAAGKVKYRGPWVYIVLRGDOPISKRAIAI 717

Qy 721 SEYDPKKHCKYDQDYYTENQVTPAVLRLTEAFGSKRDKLKYQSSKQVGLDAWIK 773

DR SMART; SM00488; POLBc; 1.

DR TIGRFAMS; TIGR00522; POL2; 1.

DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.

DR TRANSFERASE; DNA-directed DNA polymerase; 1.

DR KWD DNA-binding; Complete proteome.

FT VARIANT 532 532 R -> S (IN STRAIN GE23).

FT VARIANT 553 554 HE -> P (IN STRAIN GE23).

FT VARIANT 756 756 K -> R (IN STRAINS GE23 AND ST 855).

FT CONFLICT 263 263 V -> A (IN REF. 1).

FT CONFLICT 277 277 A -> T (IN REF. 1).

FT CONFLICT 281 281 A -> V (IN REF. 1).

FT CONFLICT 320 320 F -> S (IN REF. 1).

FT CONFLICT 339 339 Q -> H (IN REF. 1).

FT CONFLICT 359 359 R -> T (IN REF. 1).

FT CONFLICT 391 391 K -> N (IN REF. 1).

SEQUENCE 771 AA; 83496 NW; 110A87045A8A5522 CRC64;

Query Match 78.6%; Score 3164.5; Best Local Similarity 76.1%; Matches 388; Conservative 89; Mismatches 93; Indels 3; Gaps 2;

Db 1 MIFDPTDYYTDKGPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEIKAKGERHG 60

Db 1 MTDADYTDGKPKIIRIFKKGKENGFKIYDFFQPYTAYLLKDDSAIDEVKRITAHRG 60

Qy 61 KIVTRVDAVKVKKFKLGRDVEWKLIFHPDQVPAFLRGKIKRHPAVDIDYDIPPAKRY 120

Db 61 KIVTRVTEVKVQCKFLPPIEWKVLHLPQDQVPAFKRKHHPAVDIDYDIPPAKRY 120

Qy 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Db 121 LIDKGKLIIMEGDDEBLKLMADLIEFTYHEDERGKGELEMISYADEERAVITWKDIDPY 180

Qy 181 VDVSNEERMIKEPKVQIYREKDPDVLYTNGNFDLDPYLKRAKGUTVLLGRDKHEHPE 240

Db 181 VEVSSSEPKMILKLAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Qy 241 PKHRMGDSFAVJIKRHFDFPVVERTINPFTYDLEAVYTAVLTKSKLIGAEIIAI 300

Db 239 PKNCRMGSLAVBKGRHFDLFPVIRRTINPFTYDLEAVYTAFLGKSKVETVAAEA 298

Qy 301 WEFBESMCKLAAQYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 360

Db 299 WEFCKGLPRAVXYSMEDRATYELGKEFFPMAELAKLIGQSVWDVRSRSTGONLYEWYLL 358

Qy 361 RVAYERNLAPNPKPDEEFYRRLRTTLYGGYKTEPERGLWENITYLDFRCLYPSLIVTHN 420

Db 359 RKAYERNLAPNPKPDERYERLIRESEGGYKTEPERGLWGIVSDFRSRSTPSLITTHN 418

Qy 421 VSPDTLREGCKYDVAFLVGYKFCCKGKPFGETPSLIGELIMMQEIKKKKATIDIEKK 480

Db 419 VSPDTLNFENCKYDVAQGFRCKDFPGFPPSLGNLLEERQKTKRMKESDFVEKK 478

Qy 481 MLDYRORAVKLAHANSYCCYGMGPKARYTKCKEASSTWAGRYHITEMKIEEKFGKFL 540

Db 479 LLDDYRQKIKLANSYCCYGMGPKARYTKCKEASSTWAGRYDILVRELESR-GFKTL 537

Qy 541 YADTDGFATIPEKPKTIIKAKFELKFKYKINSKLPGILEEYGFYL

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE01047; ABA02362; 1; .
 EMBL; A84155; AAB87984; 1; .
 DR; PIR; S35443; S35443; .
 DR; HSSP; P56689; 1TGO.
 DR; InterPro; IPRO06172; DNA_Pol_B.
 DR; InterPro; IPRO06134; DNA_Pol_B_dom.
 DR; InterPro; IPRO06133; DNA_Pol_B_exo.
 DR; InterPro; IPRO44598; Pol12.
 Pfam; PF00136; DNA_Pol_B; 1.
 DR; PF03104; DNA_Pol_B; exo; 1.
 DR; PRINTS; PRO0106; DNAP01B.
 SMART; SM00486; POLB2; 1.
 DR; TIGRFAMS; TIGR0052; Pol12; 1.
 DR; PROSITE; PS00116; DNA_Polymerase_B; 1.
 DR; DNA-binding; Hydrolase; DNA_Polymerase; DNA replication;
 DNA-binding; Hydrolase; Endonuclease; Complete proteome.
 KW; SEQUENCE; 775 AA; 90113 MW; 566883D9F4DBD4CE CRC64;

Query Match 78.3%; Score 3151; DB 16; Length 775;
 Best Local Similarity 74.6%; Prd. No. 4.4e-164;
 Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

Qy 1 MIFDDYIINDGPIRIFKKEENGKFFKELDPHPQFPIYALKDSDAIDELKAKGRRG 60
 Db 1 MILDGLIPMEGDBELKMAFDIETFYHGDEFKGKEMIMISADEBARYTWNIDLPY 180
 Qy 61 KIVRRVDAVKVKKKFLGRVEVKLILBHPDQVPALEKQKREHPAVDIEYDPPAFRY 120
 Db 61 KIVRIVDVKVKVKEPKITVYKLYEHQDQVPTYREKVEHPAVDIEYDPPAFRY 120
 Qy 121 LIDKGGLIPMEGDBELKMAFDIETFYHGDEFKGKEMIMISADEBARYTWNIDLPY 180
 Db 121 LIDKGGLIPMEGDBELKMAFDIETLYHEGEEGKGPKIMISADEBARYTWNIDLPY 180
 Qy 181 VDVVSNEREMIKRQFVQIVREKDPVLYTNGDNFDLPLYKRAEKUGVTLGRDEKEHPE 240
 Db 181 VEVSSSEREMIKRQFVQIVREKDPVLYTNGDSDFYLAKEAKLGKLTIGRGS-E 238
 Qy 241 PTKHMGDSFAVEIKRQFEDLPVYRTRTINLPFTYLEAVIIGKTKSRGLGAEIAAAI 300
 Db 239 PTKHMGDSFAVEIKRQFEDLPVYRTRTINLPFTYLEAVIIGKTKSRGLGAEIAAAI 300
 Qy 301 WETEESMKLQAVSMEDARATYELGKEPFPMABELAKLIGOSWYDYSRSTGNYLVEWYL 360
 Db 299 WESGENLERYAKSYMEDARATYELGKEPFPMETQLSLVGQPLWDYSRSTGNYLVEWYL 358
 Qy 361 RVAYERNELAQNPKDEEYERRRRTTYGGYKREPEGLWENITYDPLCPYSPITIITH 420
 Db 359 RKAYRNBYAPKNSMEEYQRRRESTYGGFVKEPKGFLWENITYDPLCPYSPITIITH 418
 Qy 421 VSPDFTLERECKNTYDAPIVGYKPCDKDPPGFTPSLIGELITNROEIKKMKATIDPIKK 480
 Db 419 VSPDFTLERECKNTYDAPIVGYKPCDKDPPGFTPSLIGELITNROEIKKMKATIDPIKK 478
 Qy 481 MLDYCRARVLHANSYGMGYPKARYTMEKKEFKGFKVL 540
 Db 479 LDYDRKAIIKLLANSFYGTYGAKARYWCKECKESVPAWGRYTELWKEIIEKCFKVL 538
 Qy 541 YADTDFYATPGEKPKTKKKAKELKLYINSKLPGLILEYEGFVLRGEFFVAKRYAVI 600
 Db 539 YIDTDFYATPGEKPKTKKKAKELKLYINSKLPGLILEYEGFVLRGEFFVAKRYAVI 598
 Qy 601 DEEGRITTRGLEVFRDWSKIAKETQARYTLEALKEDPVEKAVETVVDVEIAYQVPL 660
 Db 599 DEEGRITTRGLEVFRDWSKIAKETQARYTLEALKEDPVEKAVETVVDVEIAYQVPL 658

RESULT 7
 DBP_PYRSID STANDARD; PRT; 1312 AA.
 ID DBP_PYRSID STANDARD; PRT; 1312 AA.
 AC Q51333; Q51335; Q51336;
 DT 01-NOV-1997 (Rel. 35, Created)
 DI 01-NOV-1997 (Rel. 35, Last sequence update)
 DR 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA Polymerase (EC 2.7.7.7) (Deep vent DNA Polymerase) [Contains:
 DE Endonuclease PI-PSPL (EC 3.1.-.-) (PSP-GDB Pol intein)].
 DR P01133; P01134; P01135; P01136; P01137; P01138; P01139; P01140; P01141; P01142; P01143; P01144; P01145; P01146; P01147; P01148; P01149; P01150; P01151; P01152; P01153; P01154; P01155; P01156; P01157; P01158; P01159; P01160; P01161; P01162; P01163; P01164; P01165; P01166; P01167; P01168; P01169; P01170; P01171; P01172; P01173; P01174; P01175; P01176; P01177; P01178; P01179; P01180; P01181; P01182; P01183; P01184; P01185; P01186; P01187; P01188; P01189; P01190; P01191; P01192; P01193; P01194; P01195; P01196; P01197; P01198; P01199; P01200; P01201; P01202; P01203; P01204; P01205; P01206; P01207; P01208; P01209; P01210; P01211; P01212; P01213; P01214; P01215; P01216; P01217; P01218; P01219; P01220; P01221; P01222; P01223; P01224; P01225; P01226; P01227; P01228; P01229; P01230; P01231; P01232; P01233; P01234; P01235; P01236; P01237; P01238; P01239; P01240; P01241; P01242; P01243; P01244; P01245; P01246; P01247; P01248; P01249; P01250; P01251; P01252; P01253; P01254; P01255; P01256; P01257; P01258; P01259; P01260; P01261; P01262; P01263; P01264; P01265; P01266; P01267; P01268; P01269; P01270; P01271; P01272; P01273; P01274; P01275; P01276; P01277; P01278; P01279; P01280; P01281; P01282; P01283; P01284; P01285; P01286; P01287; P01288; P01289; P01290; P01291; P01292; P01293; P01294; P01295; P01296; P01297; P01298; P01299; P012100; P012101; P012102; P012103; P012104; P012105; P012106; P012107; P012108; P012109; P012110; P012111; P012112; P012113; P012114; P012115; P012116; P012117; P012118; P012119; P012120; P012121; P012122; P012123; P012124; P012125; P012126; P012127; P012128; P012129; P012130; P012131; P012132; P012133; P012134; P012135; P012136; P012137; P012138; P012139; P012140; P012141; P012142; P012143; P012144; P012145; P012146; P012147; P012148; P012149; P012150; P012151; P012152; P012153; P012154; P012155; P012156; P012157; P012158; P012159; P012160; P012161; P012162; P012163; P012164; P012165; P012166; P012167; P012168; P012169; P012170; P012171; P012172; P012173; P012174; P012175; P012176; P012177; P012178; P012179; P012180; P012181; P012182; P012183; P012184; P012185; P012186; P012187; P012188; P012189; P012190; P012191; P012192; P012193; P012194; P012195; P012196; P012197; P012198; P012199; P012200; P012201; P012202; P012203; P012204; P012205; P012206; P012207; P012208; P012209; P012210; P012211; P012212; P012213; P012214; P012215; P012216; P012217; P012218; P012219; P012220; P012221; P012222; P012223; P012224; P012225; P012226; P012227; P012228; P012229; P012230; P012231; P012232; P012233; P012234; P012235; P012236; P012237; P012238; P012239; P012240; P012241; P012242; P012243; P012244; P012245; P012246; P012247; P012248; P012249; P012250; P012251; P012252; P012253; P012254; P012255; P012256; P012257; P012258; P012259; P012260; P012261; P012262; P012263; P012264; P012265; P012266; P012267; P012268; P012269; P012270; P012271; P012272; P012273; P012274; P012275; P012276; P012277; P012278; P012279; P012280; P012281; P012282; P012283; P012284; P012285; P012286; P012287; P012288; P012289; P012290; P012291; P012292; P012293; P012294; P012295; P012296; P012297; P012298; P012299; P0122100; P0122101; P0122102; P0122103; P0122104; P0122105; P0122106; P0122107; P0122108; P0122109; P0122110; P0122111; P0122112; P0122113; P0122114; P0122115; P0122116; P0122117; P0122118; P0122119; P0122120; P0122121; P0122122; P0122123; P0122124; P0122125; P0122126; P0122127; P0122128; P0122129; P0122130; P0122131; P0122132; P0122133; P0122134; P0122135; P0122136; P0122137; P0122138; P0122139; P0122140; P0122141; P0122142; P0122143; P0122144; P0122145; P0122146; P0122147; P0122148; P0122149; P0122150; P0122151; P0122152; P0122153; P0122154; P0122155; P0122156; P0122157; P0122158; P0122159; P0122160; P0122161; P0122162; P0122163; P0122164; P0122165; P0122166; P0122167; P0122168; P0122169; P0122170; P0122171; P0122172; P0122173; P0122174; P0122175; P0122176; P0122177; P0122178; P0122179; P0122180; P0122181; P0122182; P0122183; P0122184; P0122185; P0122186; P0122187; P0122188; P0122189; P0122190; P0122191; P0122192; P0122193; P0122194; P0122195; P0122196; P0122197; P0122198; P0122199; P0122200; P0122201; P0122202; P0122203; P0122204; P0122205; P0122206; P0122207; P0122208; P0122209; P0122210; P0122211; P0122212; P0122213; P0122214; P0122215; P0122216; P0122217; P0122218; P0122219; P0122220; P0122221; P0122222; P0122223; P0122224; P0122225; P0122226; P0122227; P0122228; P0122229; P01222210; P01222211; P01222212; P01222213; P01222214; P01222215; P01222216; P01222217; P01222218; P01222219; P012222100; P012222101; P012222102; P012222103; P012222104; P012222105; P012222106; P012222107; P012222108; P012222109; P012222110; P012222111; P012222112; P012222113; P012222114; P012222115; P012222116; P012222117; P012222118; P012222119; P0122221100; P0122221101; P0122221102; P0122221103; P0122221104; P0122221105; P0122221106; P0122221107; P0122221108; P0122221109; P0122221110; P0122221111; P0122221112; P0122221113; P0122221114; P0122221115; P0122221116; P0122221117; P0122221118; P0122221119; P01222211100; P01222211101; P01222211102; P01222211103; P01222211104; P01222211105; P01222211106; P01222211107; P01222211108; P01222211109; P01222211110; P01222211111; P01222211112; P01222211113; P01222211114; P01222211115; P01222211116; P01222211117; P01222211118; P01222211119; P012222111100; P012222111101; P012222111102; P012222111103; P012222111104; P012222111105; P012222111106; P012222111107; P012222111108; P012222111109; P012222111110; P012222111111; P012222111112; P012222111113; P012222111114; P012222111115; P012222111116; P012222111117; P012222111118; P012222111119; P0122221111100; P0122221111101; P0122221111102; P0122221111103; P0122221111104; P0122221111105; P0122221111106; P0122221111107; P0122221111108; P0122221111109; P0122221111110; P0122221111111; P0122221111112; P0122221111113; P0122221111114; P0122221111115; P0122221111116; P0122221111117; P0122221111118; P0122221111119; P01222211111100; P01222211111101; P01222211111102; P01222211111103; P01222211111104; P01222211111105; P01222211111106; P01222211111107; P01222211111108; P01222211111109; P01222211111110; P01222211111111; P01222211111112; P01222211111113; P01222211111114; P01222211111115; P01222211111116; P01222211111117; P01222211111118; P01222211111119; P012222111111100; P012222111111101; P012222111111102; P012222111111103; P012222111111104; P012222111111105; P012222111111106; P012222111111107; P012222111111108; P012222111111109; P012222111111110; P012222111111111; P012222111111112; P012222111111113; P012222111111114; P012222111111115; P012222111111116; P012222111111117; P012222111111118; P012222111111119; P0122221111111100; P0122221111111101; P0122221111111102; P0122221111111103; P0122221111111104; P0122221111111105; P0122221111111106; P0122221111111107; P0122221111111108; P0122221111111109; P0122221111111110; P0122221111111111; P0122221111111112; P0122221111111113; P0122221111111114; P0122221111111115; P0122221111111116; P0122221111111117; P0122221111111118; P0122221111111119; P01222211111111100; P01222211111111101; P01222211111111102; P01222211111111103; P01222211111111104; P01222211111111105; P01222211111111106; P01222211111111107; P01222211111111108; P01222211111111109; P01222211111111110; P01222211111111111; P01222211111111112; P01222211111111113; P01222211111111114; P01222211111111115; P01222211111111116; P01222211111111117; P01222211111111118; P01222211111111119; P012222111111111100; P012222111111111101; P012222111111111102; P012222111111111103; P012222111111111104; P012222111111111105; P012222111111111106; P012222111111111107; P012222111111111108; P012222111111111109; P012222111111111110; P012222111111111111; P012222111111111112; P012222111111111113; P012222111111111114; P012222111111111115; P012222111111111116; P012222111111111117; P012222111111111118; P012222111111111119; P0122221111111111100; P0122221111111111101; P0122221111111111102; P0122221111111111103; P0122221111111111104; P0122221111111111105; P0122221111111111106; P0122221111111111107; P0122221111111111108; P0122221111111111109; P0122221111111111110; P0122221111111111111; P0122221111111111112; P0122221111111111113; P0122221111111111114; P0122221111111111115; P0122221111111111116; P0122221111111111117; P0122221111111111118; P0122221111111111119; P01222211111111111100; P01222211111111111101; P01222211111111111102; P01222211111111111103; P01222211111111111104; P01222211111111111105; P01222211111111111106; P01222211111111111107; P01222211111111111108; P01222211111111111109; P01222211111111111110; P01222211111111111111; P01222211111111111112; P01222211111111111113; P01222211111111111114; P01222211111111111115; P01222211111111111116; P01222211111111111117; P01222211111111111118; P01222211111111111119; P012222111111111111100; P012222111111111111101; P012222111111111111102; P012222111111111111103; P012222111111111111104; P012222111111111111105; P012222111111111111106; P012222111111111111107; P012222111111111111108; P012222111111111111109; P012222111111111111110; P012222111111111111111; P012222111111111111112; P012222111111111111113; P012222111111111111114; P012222111111111111115; P012222111111111111116; P012222111111111111117; P012222111111111111118; P012222111111111111119; P0122221111111111111100; P0122221111111111111101; P0122221111111111111102; P0122221111111111111103; P0122221111111111111104; P0122221111111111111105; P0122221111111111111106; P0122221111111111111107; P0122221111111111111108; P0122221111111111111109; P0122221111111111111110; P0122221111111111111111; P0122221111111111111112; P0122221111111111111113; P0122221111111111111114; P0122221111111111111115; P0122221111111111111116; P0122221111111111111117; P0122221111111111111118; P0122221111111111111119; P01222211111111111111100; P01222211111111111111101; P01222211111111111111102; P01222211111111111111103; P01222211111111111111104; P01222211111111111111105; P01222211111111111111106; P01222211111111111111107; P01222211111111111111108; P01222211111111111111109; P01222211111111111111110; P01222211111111111111111; P01222211111111111111112; P01222211111111111111113; P01222211111111111111114

PRINTS: PRO0106; DNAPOLB.	Db	719	VRNGNKREYLVEFNAYRDVISMPEEBLKEWRIGTINGFRMGTVDIDEDFAKLGLYY
DR: PRO0379; INTEIN.	Qy	494	-----
SMART: SM00305; HintC; 1.	Db	779	SEGSARKWKRQNTGGWSYTVRLYNENDEVDDMEHLAKKPFKGKVKGKNEYVEPKWQAYII 838
SMART: SM00486; POLBc; 1.	Qy	494	-----
TIGR4MIS: TIGR01443; intein_Cterm; 1.	Db	839	FESICGTAAENKRVPVEIFTSSKGVRWAFLGGYFIGGDVHPSKRYVRLSTKSSELLVNGLV 898
TIGR4MIS: TIGR01445; intein_Nterm; 1.	Qy	494	-----
DR: TIGR0592; pol2; 1.	Db	899	LLINSLGSAIKLGDSGYRVVNEELKFETYRKVKVQVHSHIVPKDILKETFGKVYQK 958
DR: PROSITE: PS00116; DNA POLYMERASE_B; 1.	Qy	494	-----
DR: PROSITE: PS50819; INTEIN_CTER; 1.	Db	1019	LAGPGFLYAHNSYYGGYAKARWYCKECAESTWAGREYIEFVKELEEKFGKVLYID 1078
DR: PROSITE: PS50817; INTEIN_NTER; 1.	Qy	494	-----
DR: PROSITE: PS0818; INtron homing.	Db	959	NISYKFERLVENGKLDREKAKEAKRIEWLNGDIVLDRVEEIKREYDGYVYDLSVDEDENF 1018
KW: DNA-directed DNA polymerase; DNA replication; DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage; Protein Splicing;	Qy	494	-----
CHAIN 1 492 1029 DNA POLYMERASE PI-PSPI.	Db	1079	TDGFYATIPEGKPTIKKAKAKETFLKINSKLPGLLELEYEGFLYLRGFFVAKCRYAVDEE 603
CHAIN 1 493 1312 DNA POLYMERASE, 2ND PART.	Qy	544	TDGFYATIPEGKPTIKKAKAKETFLKINSKLPGLLELEYEGFLYLRGFFVAKCRYAVDEE 603
SEQUENCE 1 1312 AA; 862518805641D26A CRCC4;	Db	1079	TDGFYATIPEGKPTIKKAKAKETFLKINSKLPGLLELEYEGFLYLRGFFVAKCRYAVDEE 603
Query Match Score 2934.5; DB: 1; Length 1312;	Qy	1079	TDGFYATIPEGKPTIKKAKAKETFLKINSKLPGLLELEYEGFLYLRGFFVAKCRYAVDEE 603
Best Local Similarity 45.6%; Pred. No. 5e-152; Gaps 3;	Db	1119	TDGFYATIPEGKPTIKKAKAKETFLKINSKLPGLLELEYEGFLYLRGFFVAKCRYAVDEE 603
Matches 597; Conservative 86; Mismatches 87; Indels 539; Gaps 3;	Qy	604	GRITTRGLEVVRDWSLELAKETQAKVLETAIKLKDSEVKAELYKQVPLEKLI 663
1 MIFDTDTYITKGGKPITRPKKENGEGPKLIDDPHQFQPYIYALKDOSAIDDEKAIGKPRIG 60	Db	1119	GRITTRGLEVVRDWSLELAKETQAKVLETAIKLKDSEVKAELYKQVPLEKLI 663
1 MILDADITEGKPITRPKKENGEGPKLIDDPHQFQPYIYALKDOSQIDEVRKITAERIG 60	Qy	664	VIEHQITDSEKKAIGPHVIAIKRLAKGKIVPRGTTISIYVRGSGSKISDRVILSEY 723
61 KIVRVDAYKVKKKFLGRDVWKLIFPEHPODPVAPLCKCREHPAVIDYEDIPFAKRY 120	Db	1119	VIEHQITDSEKKAIGPHVIAIKRLAKGKIVPRGTTISIYVRGSGSKISDRVILSEY 723
61 KIVRVDAYKVKKKFLGRDVWKLIFPEHPODPVAPLCKCREHSAVIDFEYDIPFAKRY 120	Qy	724	DPKKHKYIDPDIYENQVPAVIRILEAFGYRZEDLKKQSSKGVDLAMLI 772
121 LIDKGKJLPMEGDDEEKLKMAFDIETFYHEDFGKGEIIMISYADEBEEARVITWNKIDPY 180	Db	1119	DPKKHKYIDPDIYENQVPAVIRILEAFGYRZEDLKKQSSKGVDLAMLI 772
121 LIDKGKJLPMEGDDEEKLKMAFDIETLYHEGEEBAKSPIMIISYADEBEEARVITWNKIDPY 180	Qy	1259	DIRKHKYDAEYIYENQVPAVIRILEAFGYRZEDLKKQSSKGVDLAMLI 1307
Db	1259	DIRKHKYDAEYIYENQVPAVIRILEAFGYRZEDLKKQSSKGVDLAMLI 1307	
RESULT 8			
DPOI_PYRHO			
ID_DPOI_PYRHO			
AC_05910;			
DT 15-DEC-1998 (Rel. 37, Created)			
DT 15-DEC-1998 (Rel. 37, Last sequence update)			
DT 28-DEB-2003 (Rel. 41, Last annotation update)			
DE DNA polymerase (EC 2.7.7.7) [Contains: Pho pol integrin; Pho pol integrin].			
DE integrin].			
GN POL OR PH1947 OR PHBT047.			
OS Pyrococcus horikoshii.			
OC Archaea: Euryarchaeota;			
NCBI_TaxID=53953;			
[1] N.C. -			
RN SEQUENCE FROM N.A.			
RC STRAIN=OT13;			
RX MEDLINE=8344137; PubMed=9679194;			
RA KAWABRAYASI Y., Swada M., Horikawa H., Haino Y., Hino Y.,			
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,			
RA Sakai T., Otsuka S., Nakazawa H., Takemoto M., Ohnishi Y.,			
RA Furukashiba T., Takemoto T., Kudo H., Yamazaki J., Kubishi N., Oguchi A.,			
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.M., Horikoshi K.,			
RA Masuchi Y., Shizuya H., Kuroki H.,			
RT "Complete sequence and gene organization of the genome of a hyper-			
RT thermophilic archaeabacterium, Pyrococcus horikoshii OT3.",			
RL DNA Res. 5:55-76 (1998).			
CC - CATALYTIC ACTIVITY. N deoxyribonucleoside triphosphate = N diphosphate			
CC + (DNA) (N).			
CC - PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN) FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).			
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).			
CC CONSTITUTE BELONGS TO THE DNA POLYMERASE TYPE B FAMILY.			
CC			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AP000007; BAA31074.1; -.

DR; PIR; C71210; C71210.

DR; HSSP; P56689; 1TGO.

DR; InterPro; IPRO01172; DNA_Pol_B.

DR; InterPro; IPRO01134; DNA_Pol_B_dcm.

DR; InterPro; IPRO01133; DNA_Pol_B_exo.

DR; InterPro; IPRO03586; Hedgehog_FntN.

DR; InterPro; IPRO03587; Hedgehog_FntN.

DR; InterPro; IPRO06141; Intein.

DR; InterPro; IPRO06142; INTEIN.

DR; InterPro; IPRO0402; Intein_endonuc.

DR; InterPro; IPRO04578; Pol2.

DR; Pfam; PF00136; DNA_Pol_B_1.

DR; Pfam; PF03104; DNA_Pol_B_exo; 1.

DR; PRINTS; PR00106; DNAPOLB.

DR; SMART; SM00305; HintC; 1.

DR; SMART; SM00306; HintN; 1.

DR; SMART; SM00486; PolBc; 1.

DR; TIGRFAMS; TIGR01443; intein_Cterm; 1.

DR; TIGRFAMS; TIGR01445; intein_Nterm; 1.

DR; TIGRFAMS; TIGR0592; Pol12_1.

DR; PROSITE; PSS0016; DNA POLYMERASE_B; 1.

DR; PROSITE; PSS0018; INTEIN_C_TER; 1.

DR; PROSITE; PSS0819; INTEIN-ENDONUCLEASE; 1.

DR; PROSITE; PSS0818; INTEIN_N_TER; 1.

KW; Transferase; DNA-directed DNA polymerase; DNA replication; DNA-binding; Autocatalytic cleavage; Protein splicing;

KW; DNA-binding Proteome.

KW; Complete Proteome.

COMPATIBILITY: 100%.

CHAIN: 1. 492 DNA POLYMERASE, 1ST PART (POTENTIAL).

FT CHAIN 493 952 PHO POLYMERASE, 2ND PART (POTENTIAL).

FT CHAIN 953 1235 DNA POLYMERASE, 2ND PART (POTENTIAL).

FT SEQUENCE 1235 AA; 143086 MW; 73CC7AA14873C84 CRC64;

Score 71.7%; Score 2886; DB 1; Length 1235;

Best Local Similarity 46.5%; Prod. No. 2e-149; Gaps 5;

Matches 574; Conservative 107; Mismatches 89; Indels 464; Gaps 5;

Query Match 1. MIFDDDYTKDGPPIRKKENGEPKIEFLDPHPQPIYALLKDDSDIDEIKAIKGHRG 60

1. MILDADYYTEDGPPIRKKENGEPKIEFLDPHPQPIYALLRDDSDIDEIKKITAQRHG 60

61 KIVRVVDAVKKKKFLGROEVNKKLIFEPDQVPAKIRKIREPAVVDIYEDIPPAKRY 120

61. KIVRVVETEKIQKRFGLGRPIEVNKKLLEHFDQVPAIRKIREPAVVDIYEDIPPAKRY 120

QY 121 LIDKGLIPMEGDEELKLMADFIEFYHEGDEFGKGEITIMISTADEEARVITWNIDLPY 180

121 LIDKGLIPMEGNEKLTFLADTLYHEGEEFKGPVIMISTADEEAKVITWNIDLPY 180

181 VDVVNERENIKRPFQIVREKDPVYDPLITNGDNFLPLTILKRAEKGTLILGRDRHPE 240

181 VEVVSEERENIKRILRVIKRDQDVTIITNGDNFLPFLKRAEKGTLIKLGRD--NSE 238

241 PKHMRMGDSPAVEIKRGIHEDLPYTRVTLNLPYTHEAVLGTIKTSKLGAEIAAI 300

239 PKMQKNGDSLAVEIKRHIIDLPYTRTINLPYTHEAVLGTIKTSKLGAEIAAI 298

301 WETEESMKKLAQYSMDARATYELGEKEFFPMEALAKLKGQSWSYDVSRSSTGNLVEVNTL 360

299 WETGEGLERVAKYMSMDAKVYTYELGEKEFFPMEALAKLKGQSWSYDVSRSSTGNLVEVNTL 358

361 RVAYENELAANKPDEBEYRRLRSTYLGWVKEPERGLWNTYDPERCLYPSI1IITHN 420

Db 359 RKAYERNELAANKPDEKEYERRLRLRSTYLGWVKEPERGLWNTYDPERCLYPSI1IITHN 418

QY 421 VSPDTLREBECKNDVAP1VGYKEKCDPFFGFIISLIGELITMRQEIKKMKATIDPIEKK 480

Db 419 VSPDTLREBECKNDVAP1VGYKEKCDPFFGFIISLIGELITMRQEIKKMKATIDPIEKK 478

QY 481 MLDYRCRAVL----- 491

Db 479 LLDYRORAIKILANSILPDEWLPIVENEKVRFYKIGDFIDREIEBENAERVKRDGETEEL 538

QY 492 ----- 491

Db 539 VRDLKALSFNRETKSKESELKVKALLHRYSGKVYSIKLKGRRKITSGHSLSVYNGKRL 598

QY 492 ----- 491

Db 599 VKYRGDELKPGDILVUVPGRLKLPERSKQVINVELLKLPEEETSNIVMIPVKGRKNFFK 658

QY 492 ----- 491

Db 659 GMLKLTLYWIFGEGERBRTAGYLKHLERLGYVVKURRGCEVLDNESSLKRYKLYTSLIKN 718

QY 492 ----- 491

Db 719 LKYNQNSRAYMVEENSLRDVYSLMP1EELKEMWIGEPRGPKIGTIDVDDDSPAKLLGYI 778

QY 492 ----- 493

Db 779 SSSDVEKDVKPHSKDQVNLIEDIAKLAELKFQGKVRGRGYIEVSGSKISHAIFRVLAEGKR 838

QY 494 ----- 493

Db 839 IPEPIFTSPMDIKVAPLKGNGNAEELTFSTKSELVNLQILLNNSIGVSDIKIHEKGV 898

QY 494 ----- 493

Db 899 YRYTINKKESSNNDIVLDSVESIEVEKVEGYVYDLSVEDENFLVFGFLYAHNSYGY 958

QY 501 GYPRKARYNSKECAESVTAWGHRHYIEMTKEEKFGFKVLYADTDDGTYATIFGEKP-ETI 559

Db 959 GYAKARWCKECAESVTAWGQYDLYTRELAR-GRFKVLYIDTDDGTYATIFGVKWEV 1017

QY 560 KCKZKEFELKIVNSKLPGLLEBEFGYLRGFFVAKKRYAVIDBEGRITTRGEBVVRDWS 619

Db 1018 KRAZALEFVDYDINSKLPGVLEBEFGYARGFFETKRYALDEBEGKIVTRGEBVVRDWS 1077

QY 620 EIAKETQATKLEALKEDSVEKAEVYDVKVBEIAKQVPLFKLVTHEQITDSEYKAI 679

Db 1078 EIAKETQATKLEALKHGVVEAKVKEVTDVTEKLNTVEPPVLEVIEQTRPNEKAI 1137

QY 680 GPHAAIAKLAAGKJVKRPGTISIYVLRGSKGKISDRVILLSEYDPKKHHKTYDPPYYENQ 739

Db 1138 GPHAAVAKLIMARGSKVKGPMVQYIVLRGDGPKSKRAISIEFDPRKHKTDAAEYYENQ 1197

QY 740 VLPAYTRILLEAFGGTRKDULYQSSKQVGILDANIK 773

Db 1198 VLPAVERILKAFGTYKREDLWRQKTKQVGLGAWIK 1231

RESULT 9

DPOL THEFM STANDARD PRT; 1523 AA.

ID DPOL THEFM ID P74919; AC P74919; DT 15-DEC-1998 (Rel. 3.7, Created)

DT 15-DEC-1998 (Rel. 3.7, Last sequence update)

DT 28-DEC-2003 (Rel. 4.1, Last annotation update)

DE DNA Polymerase (EC 2.7.7.7) (Pol_Tfu) (Contains: Endonuclease PI-Tfu; DE (Tfu_Pol-2) (Intein); Endonuclease PI-Tfu) (EC 3.1.-.-)

DE (Tfu_Pol-2) (Intein); Endonuclease PI-Tfu) (EC 3.1.-.-)

GN PCL.

OS Thermococcus funicolans

OC Archaea; Buryarchaeota; Thermococci; Thermococcaceae;

OC Thermococcus.

NCBI_TAXID=46540;
 RN [1] SEQUENCE FROM N. A.
 STRAIN=ST557;
 RC Cabon M., Querellou J.; Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RA [2] CHARACTERS OF INTEINS.
 RC MEDLINE=20112788; PubMed=10644683;
 RX [1] "Inteins of *Thermococcus fumicola* and *M. marcusii* are endonucleases with distinct enzymatic behaviors.", J. Biol. Chem. 275:233-244 (2000).
 CC -!- FUNCTION: PI-Tfull recognizes and cleaves a minimal sequence of 16 base pairs (bp) on supercoiled DNA with either Mn²⁺ or Mg²⁺ as cofactor. It cleaves linear DNA only with Mn²⁺ and requires a 19 bp minimal recognition sequence. The optimal temperature for activity is 70 degrees Celsius.
 CC -!- PI-Tfull is a highly active homing endonuclease using Mg²⁺ as cofactor. Its minimal recognition and cleavage site is 21 bp long either on linear or circular DNA substrates. Its endonuclease activity is strongly inhibited by the 3' cleavage product, which remains bound to the enzyme after the cleavage reaction. The optimal temperature for activity is 70 degrees Celsius.
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + (DNA) (N).
 CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES A POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION (INTEINS) FOLLOWED BY PEPTIDE LIGATION.
 CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
 CC -!- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING ENDONUCLEASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the ENBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z69882; CAA9378.1; -.
 DR HSSP; P56699; 1TGO.
 DR REBASE; 4500; PI-Tfull.
 DR REBASE; 4501; PI-Tfull.
 DR InterPro; IPR006172; DNA_Pol_B.
 DR InterPro; IPR006134; DNA_Pol_B_exo.
 DR InterPro; IPR006133; DNA_Pol_B_dom.
 DR InterPro; IPR003586; Hedgehog_HintN.
 DR InterPro; IPR003587; Hedgehog_HintN.
 DR InterPro; IPR006141; Intein.
 DR InterPro; IPR006142; Intein.
 DR InterPro; IPR004042; Intein_endonuc.
 DR InterPro; IPR004578; Pol12.
 DR Pfam; PF001336; DNA_Pol_B.
 DR Pfam; PF0104; DNA_Pol_B_exo.
 DR PRINTS; PRO0379; INTEIN.
 DR SMART; SM00305; HintC.
 DR SMART; SM00306; HintN.
 DR SMART; SM00486; PolBc.
 DR TIGRFAMS; TIGR01443; intein_Cterm; 2.
 DR TIGRFAMS; TIGR01445; intein_Nterm; 2.
 DR PROSITE; PS500116; DNA_POLYMERASE_B; FALSE_NEG.
 DR PROSITE; PS500818; INTEIN_CTER; 2.
 DR PROSITE; PS500819; INTEIN_NTER; 2.
 DR Transferase; PS500817; INTEIN_NTER; 2.
 KW DNA-binding; Hydrolase; Endonuclease; DNA replicase; DNA-directed DNA polymerase; DNA replication.
 KW Protein_splicing; Intron_splicing; Magnesium; Manganese.
 CHAIN 1; 406 DNA POLYMERASE, 1ST PART.

FT CHAIN 407 766 ENDONUCLEASE PI-Tfull.
 FT CHAIN 767 900 DNA POLYMERASE, 2ND PART.
 FT CHAIN 901 1282 ENDONUCLEASE PI-Tfull.
 FT CHAIN 1283 1523 DNA POLYMERASE, 3RD PART.
 SQ SEQUENCE 1523 AA; 175917 MW; 7A2C8236BF2B5F CRC64;
 Query Match 69.8%; Score 2809; DB 1; Length 1523;
 Best Local Similarity 39.0%; Prd. No. 4.1e-15;
 Matches 593; Conservative 86; Mismatches 91; Indels 752; Gaps 4;
 Qy 1 MIFDTDXITKDGKPIIRFKENGEPEKIELDPFPQPTYALLKDDSAIDEIKAIKGGRHG 60
 Db 1 MILDQDTYEDGPVTRFKENGEPEKIEDDPEPTYALLKDDSAIDEIKAIKGGRHG 60
 Qy 61 KIVYVDAVYKUKKKPLGRDVEWTKLIFPHQDVPALARQKIRHHPAVDIFYEDIPFAKRY 120
 Db 61 TTIVRVRAGKVKKKFGRPIEWKLYSTHPQDVPALARQKIRHHPAVDIFYEDIPFAKRY 120
 Qy 121 LIDKGKLIPIEGDDEBKLMAPDIDTFBEGDDEPGKGEIMISVADEBPARVTKWKLIDLPY 180
 Db 121 LIDKGKLIPIEGDDEBKLMAPDIDTFBEGDDEPGKGEIMISVADEBPARVTKWKLIDLPY 180
 Qy 181 VDVVSNEBEMIKRFVQIVREKDPDVLITYNGDNFDLPLVLTGKVLGIGDEKEHPE 240
 Db 181 VDVVSNEBEMIKRFVQIVREKDPDVLITYNGDNFDLPLVLTGKVLGIGDEKEHPE 238
 CC +!- 238
 Qy 241 PKTHRGDPSFAYBTKGRTHFDLFPVYRTRINLPTYLEAVYAVLGTKTSKLGAEEETAAI 300
 Db 239 PKTQRMGDRFAVEVKGRTHFDLFPVYRTRINLPTYLEAVYAIFFQPKEVYAAETAAQ 298
 Qy 301 WETEESMRKLAQYSMEDARATEYLGKFFPMEAELAKLIGOSWMDYRSSTGNYLWYLL 360
 Db 299 WETEGGLERVATSMEDAKVTELGKFFPMEAQLSSLVQFWDYRSSTGNYLWYLL 358
 Qy 361 RVAAYERNELAPNPKPDEBEYRRRLRRTTLYGGVYKEPBRGLMENITYLDFRC----- 410
 Db 359 RKAYERNELAPNPKPSGRELERR-RGGTAGGYKEPGRGLWNTAYLDFRCHPADTRVVK 417
 CC -!- 417
 CC -!- 410
 CC -!- 418 GKVGVVISEVRGDDYVGLDGWQKVQRWVEXDYEGBLVNINGLKC7PNHKLPVVRETERQ 477
 CC -!- 411
 CC -!- 410
 DR 478 TAIRDSLAKSFLTKVKGKLTTPLEKIGKIEREDVPEEILKGELLAGILAEGLLRLK 537
 DR 411 -----
 DR 538 DVEYFDSSRGKERVSHQYRVEITVGAQEDFQRRIYIFERLFGVTPSFRKNTNTAIF 597
 DR 411 -----
 DR 598 KVAKCEVYLVRVREIMDGCIENLHAPSVLRGFPBGDSVNVKVRKTVVVNQGTNNEWKTEVVS 657
 DR 411 -----
 DR 658 KLLNKG1SPHRYTYTYDTERKTMTHILEIAGRDLILFQTTIVGFTERNMALKEAR 717
 DR 411 -----
 DR 718 NREVNRLLENNAAFTLADFTAKTEYYKGKVYDLTBLGTPYFANGLTHNLYPSLISHN 777
 DR 421 VSPDTLREGCKNYDAPIVGYKFCKDFFGPIPSLIGELITMRQEIKKNTKATPIEKK 480
 DR 778 VSPDTLNRGGCCEYDAPQSHRCFDFFGPIPSLIGELDUDERQKVKHAKATVDEIEKK 837
 DR 481 MLDYRQAVKUHLANSVYGYMOPKAHWYSKCAESVTAWGRHYIEMTKELEKGGFKVL 540
 DR 838 LDYRQRAIKILANSVYGYGKAWYCRECAESTAWGRQYIETMREIEEKGFKVL 897
 DR 541 YAD-----
 DR 898 YADSVTGDTEVTRRGRRIEFVPIEKLFERVDRHVGKEYCULGVGEALTLDNGRLVNK 957

Nucleic Acids Res. 26:4409-4412 (1998).
 -1- FUNCTION: INTEIN ENCODED. ENDONUCLEOTIC
 -1- INTEIN MOBILITY BY SITE-SPECIFIC RE-
 ENDONUCLEASE CLEAVAGE AT THE "HOMIN"
 INTEIN. PI-PROT. RECOGNIZES 5'-GATT-
 RECOGNIZES 5'-CAGCTTACGGTAC-3'.
 -1- CATALYTIC ACTIVITY: N deoxynucleos-
 + (DNA) (N).
 -1- PTM: THIS PROTEIN UNDERGOES A PROT-
 A. POST TRANSLATIONAL EXCISION OF THE
 PROTEIN BY PEPTIDE LIGATION (POTE).
 -1- SIMILARITY: BELONGS TO THE DNA POL-
 -1- SIMILARITY: IN THE INTEIN SECTION;
 ENDONUCLEASE FAMILY.
 -1-
 EMBL: D20671; BAA6142.2; -.
 PDB: 1GCX; 28-JAN-03.
 REBASE; 3792; PI-PK01.
 REBASE; 3793; PI-PK01.
 InterPro; IPR006122; DNA_pol_B.
 InterPro; IPR006144; DNA_pol_B_exo.
 InterPro; IPR006133; DNA_pol_B_exo.
 InterPro; IPR003365; Hedgehog_HintC.
 InterPro; IPR003587; Hfq_HintN.
 InterPro; IPR006141; Intein_HintN.
 InterPro; IPR006132; INTEIN.
 InterPro; IPR004042; Intein_endonuc.
 InterPro; IPR004578; Pol12.
 Pfam; PF00136; DNA_pol_B; 3.
 Pfam; PF03104; DNA_pol_B_exo; 1.
 PRINTS; PRO0319; INTEIN.
 SMART; SM00305; HintC; 2.
 SMART; SM00306; HintN; 2.
 SMART; SM00486; POLBC; 1.
 TIGRFAMS; TIGR01453; intein_Cterm; 2.
 TIGRFAMS; TIGR00522; Pol12; 2.
 PROSITE; PS00116; DNA_BOLYMERASE_B; 1.
 PROSITE; PS50018; INTEIN_CTER; 2.
 PROSITE; PS50019; INTEIN_ENDONUCLEASE;
 PROSITE; PS50017; INTEIN_NTER; 2.
 Transferase; DNA-directed DNA polymerase
 DNA-Binding; Hydrolase; Endonuclease;
 Protein_splicing; Intein_homing; 3D_st-
 CHAIN 1 406 DNA POLYM
 CHAIN 1 766 ENDONUCLE
 CHAIN 407 851 DNA POLYM
 CHAIN 767 1388 ENDONUCLE
 CHAIN 852 1671 DNA POLYM
 CHAIN 1389 193490 MW; 4A17P
 SEQUENCE 1671 AA; 193490 MW; 4A17P
 Query Match 69.1% Score 278
 Best Local Similarity 36.0% Prod. No.
 Matches 602; Conservative 82; Mismat.
 1 MIFDDYIYKDGPKIPIRKKEKGEKIE
 1 MILDIDYIYDGPVIRKKEKGEKIE
 1 MILDIDYIYDGPVIRKKEKGEKIE
 61 KIVRVDANKVKKKKFLGDRIVEWKLIFH
 61 TIVVYTKRVEKVKQKFLGRVEVWKLFLYHG
 121 LIDKGLIPMEGDEBLKUMAFD1STYHG

RL	Nucleic Acids Res.	26:4409-4412 (1998).
CC	-1- FUNCTION: INTein ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE	
CC	INTein MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY	
CC	ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE	
CC	INTein. PT ₁ -PKO1 RECOGNIZES 5'-GATTAGATCCGTGACC-3' AND PT ₁ -SKO1	
CC	RECOGNIZES 5'-CAGCTACTACGGTAC-3'. BOTH ARE THERMOSTABLE.	
CC	-1- CATALYTIC ACTIVITY: N deoxyribonucleoside triphosphate = N diphosphate	
CC	+ {DNA} (N).	
CC	-1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES	
CC	A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)	
CC	FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).	
CC	-1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.	
CC	-1- SIMILARITY: IN THE INTEIN SECTION, BELONGS TO THE HOMING	
CC	ENDONUCLEASE FAMILY.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
DR	EMBL; D29671; BAA06142; 2; -.	
DR	PDB; 1GCX; 2B -PAN-03.	
DR	REBASE; 3792; PI-ProT.	
DR	REBASE; 3793; PI-ProT.	
DR	InteroPro; IPR00172; DNA_pol_B.	
DR	InteroPro; IPR006134; DNA_pol_B_dom.	
DR	InteroPro; IPR006133; DNA_pol_B_exo.	
DR	InteroPro; IPR003586; Hedgehog_Hntc.	
DR	InteroPro; IPR003587; Hedgehog_hirln.	
DR	InteroPro; IPR006141; Hedgehog_hirln.	
DR	InteroPro; IPR006142; INTEIN.	
DR	InteroPro; IPR004042; INTEIN_endonuc.	
DR	InteroPro; IPR004578; Pol12.	
DR	PF00136; DNA_Pol_B; 3; 1.	
DR	PF001104; DNA_Pol_B_exo; 1.	
DR	PRINTS; PR00379; INTEIN.	
DR	SMART; SM00305; Hntc.	
DR	SMART; SM00306; HntcN.	
DR	SMART; SM00486; POLBC.	
DR	TIGRFAMS; TIGR01443; intein_Cterm; 2.	
DR	TIGRFAMS; TIGR01445; intein_Nterm; 2.	
DR	TIGRFAMS; TIGR00592; pol2; 2.	
DR	PROSITE; PS00116; DNA_Polymerase_B; 1.	
DR	PROSITE; PS50818; INTEIN_C_TER; 2.	
DR	PROSITE; PS50819; INTEIN_N_TER; 2.	
KW	Transferase; DNA-directed DNA polymerase; DNA replication;	
KW	Protein splicing; Endonuclease; Autocatalytic cleavage;	
KW	Protein binding; Hydrolase; Endonuclease; 3D-structure.	
FT	CHAIN 1 406 DNA POLYMERASE; 1ST PART.	
FT	CHAIN 407 766 ENDONUCLEASE PI-PRO I.	
FT	CHAIN 767 851 DNA POLYMERASE; 2ND PART.	
FT	CHAIN 852 1388 ENDONUCLEASE PI-PRO II.	
FT	CHAIN 1389 1671 DNA POLYMERASE; 3RD PART.	
SQ	SEQUENCE 1671 AA; 193490 MW; 4A17F1C8120EB455 CRC64;	
Query Match	69.1%	Score 2782; DB 1; Length 1671;
Best Local Similarity	36.0%	Pred. No. 1..3e-13;
Matches	602;	Mismatches 86; Indels 900; Gaps 4; A
Qy	1 MIFDDYITDGPVKIRPKKENGEKFKEIQLDHPQDVPALGKIRHPAVIDYDIDPFAKRY 120	
Db	1 MILDFTYITDGPVKIRPKKENGEKFKEIQLDHPQDVPALGKIRHPAVIDYDIDPFAKRY 60	
Qy	61 KIVVVDAVKVKKGKFLGRDVEVKMFLIFEPQIYALKKDSADEIKAIKGERRG 60	
Db	61 TVVTVKRVVKQFLGRDVEVKMFLIFEPQIYALKKDSADEIKAIKGERRG 60	
Qy	121 LIDKGLJLPMGDEIQLKWDADTIEPQHGEDEFSKGEITIMSVADEBEBARVTKNQDLPY 120	

Db	121	LIDKGLVPMEGDBELKMLAFDIEFLYHEGEEPAEGPILLIMISVADEECARVITWKNDLPLV	180	Qy	494	-----	
Qy	181	VDYVSNEREMIKRFVOLYREKOPDVLLTYNGDNFDLPLYLIKRAEKGVTLIGRKEHPE	240	Db	1258	LLINSGVGSAVKLGHDGGYVYVINEELPFVKLDKRNAYAYSHVIPKREVLSVFGKVLFQK	1317
Db	181	VDYSTEEMIKRFVFLVREKOPDVLLTYNGDNFDAYLKRKCERGGINFALGRDSE-E	238	Qy	494	-----	
Qy	241	PKTHRMGGSFAVYBTKGRHFDLFPVYRTRINLPTYLBAYVAGLTKTSKLGAEBIAI	300	Db	1318	NVSPOTERKRNVEDGRDLPKAQRLSLWLBEGVVLDLDRVESVDYDGYVYDLSVEDDNNF	1377
Db	239	PKIQRMGDRFAVYBVKGRHFDLFPVYRTRINLPTYLBAYVAGLTKTSKLGAEBIAI	298	Qy	494	-----	
Qy	301	WEFEESKKLQYSMDSBARTAYLKEFPEPMMAELAKUQGQWMDSSRSGNLYEWLL	360	Db	1378	LGVFSLVYAHNSYYGGYARAWYCKECAESTAAGREYIMTKEIIEKYGFKVYAD	543
Db	299	WEGENLNRVARYSMEDKRVYBGLKEFPEPMMAELAKUQGQWMDSSRSGNLYEWLL	358	Qy	494	-----	
Qy	361	RVAYERNLAPANPKPDEEYRRLRTTLLGGTYKEPERGLWENITYLDFRC	410	Db	1438	TDGFPATIGKEPKTIIKKAKKEFLKYGINSKLPGLELEYEGFYLRGFFVAKRKYAVIDEE	1497
Db	359	RVAYERNLAPANPKPDEKLLR-ROSTEGGYKTYKEPERGLWENITYLDFRC	417	Qy	604	GRITTRGLEVVRDWSSELAKEQARVLEALKEDSVEKAEIVRDVVEIAYKQVPLKEL	663
Qy	411	-----	410	Db	1498	GKIRTRGLEVVRDWSSELAKEQARVLEALKDGVDEKAVRIVKETEKLSKYEVPPBK	1557
Db	418	GKGTININSEVQEGDYVLGIDGMQRVRKWEYDYGELVNINGLKCTPNPKLPUVTKNBRQ	477	Qy	664	VIHEQITKDSEYKAIGPHVIAKRLAARGKIVRGTIIISYVLRGSGKISDRVILSEY	723
Qy	411	-----	410	Db	1558	VIHQQTIRLKDVKATGPHVAVAKRLARGVKIRGTVISYVLRGSGRIGDRAIPFDEF	1617
Db	478	TRIDSLAKSPLTKVKCQKIIITPLFYIGRATSENIPEEVTLKGELAGILLAEGTLRK	537	Qy	724	DPKXKYDPPYYENQVLPAVLRLTEAFGYRKEDLKYQSSKOVGLDAWK	773
Qy	411	-----	410	Db	1618	DPTXKHYDAEYYENQVLPAVERLRLTAFGYRKEDLQKTRVGLSAMLK	1667
Db	538	DVEYFDSSRKKREISHQYRVEITIGDDEEFDRITYIFERLFGITPSISERKGNTAVL	597	-----	RESULT 11.		
Qy	411	-----	410	DPOL THEG8	STANDARD;	ERT;	1699 AA.
Db	598	KVAKKNNYLVKVBIMDNIBSLHAPSVLRGFFGDGSYNVRVRSIVATQGTNWKIKLVS	657	ID	DPOL THEG8		
Qy	411	-----	410	AC	Q9H864;		
Db	658	KILSQLGIPHTOTYQYQENGKDRSRYLIEITGKDGLLFTQLTGTSERKNALLNTKAS	717	DT	16-OCT-2001 (Rel. 40, Created)		
Qy	411	-----	410	DT	16-OCT-2001 (Rel. 40, Last sequence update)		
Db	718	QRENNNTLENGFYRLSEFNVSTTEYEGKRYDITLEGTPYYFANGILTHNSLYPSIITEN	777	DB	28-FEB-2003 (Rel. 41, Last annotation update)		
Qy	421	VSPDTLLEBGKRNQYDAIIVGYKFCKDPLGFTPSIOLBLITMREQIKKOMKATIDIERK	480	DE	DNA Polymerase (BC 2.7.7.7) [Contains: Endonuclease PI-TspGE8I		
Db	778	VSPDTLNEREGCKBVDQVGHFCKDPLGFTPSIOLBLITMREQIKKOMKATIDIERK	837	DE	[DE 3.1.---] (TSP-GE8 pol-1 intein); Endonuclease PI-TspGE8II		
Qy	481	MLDYRORAVKLHA-----	493	GN	POL OR POL-1.		
Db	838	LLDYRQAKILANSILPEWPLVLEGEVHFRIGELIDRMMEENAGVKREGETEVLE	897	OS	Thermococcus sp. (strain GBE).		
Qy	494	-----	493	OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococaceae;		
Db	898	VSGLEVPSFNRRNPKAELKVKALIRHDYSGRKVTIRLSGRKITSGHLSFSVRNGEL	957	OC	Thermococcus		
Qy	494	-----	493	RN	[NCB_ TaxID:105583; [1]]		
Db	958	VEVTGDELKPGDIAVAPRRLPELPERNHVNLVELLIGTPEEETLIVDWSLKVNTRLYALVEN	1077	RP	SEQUENCE FROM N. A.		
Qy	494	-----	493	RA	Querellou J.J.E., Cambon M.A., Lescoueur F., Barbier G.;		
Db	1018	GMRLTRWTFGEERPRARRYLRLHELDGYVRLKKGYEVILDWDLSLKNTVRLYALVEN	1137	RT	"Thermococcales taxonomy and phylogeny based on the comparative use of		
Qy	494	-----	493	RT	16S rDNA, 16S-23S rDNA intergenic spacer and Family B DNA polymerase		
Db	1078	VRYNGKREYLVEFNISRDVGIMPLKELKEMKIGTLNGFMRKLIEVEDSLAKLGLYY	1137	RT	genes";		
Qy	494	-----	493	RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.		
Db	1136	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	-!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE		
Qy	494	-----	493	CC	-!- EXHIBITS 3, TO 5, EXONUCLEASE ACTIVITY (BY SIMILARITY).		
Db	1194	PFNMGVTLAENKRIPEFTVFTSPKGVRNFILEGYYF1GDDVHPNKRDLRSLTRSELANQLV	1257	CC	-!- FUNCTION: PI-TSPGE8I AND PI-TSPGE8II ARE ENDONUCLEASES		
Qy	494	-----	493	CC	-!- (POTENTIAL).		
Db	1196	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	-!- CATALYTIC ACTIVITY: N deoxyribofuranoside triphosphate		
Qy	494	-----	493	CC	-!- + [DNA] (N).		
Db	1198	PFNMGVTLAENKRIPEFTVFTSPKGVRNFILEGYYF1GDDVHPNKRDLRSLTRSELANQLV	1257	CC	-!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES		
Qy	494	-----	493	CC	CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)		
Db	1199	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	CC FOLLOWED BY PEPTIDE LIGATION.		
Qy	494	-----	493	CC	CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.		
Db	1199	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	CC -!- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING		
Qy	494	-----	493	CC	CC ENDONUCLEASE FAMILY.		
Db	1199	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration		
Qy	494	-----	493	CC	CC between the Swiss Institute of Bioinformatics and the EMBL outstation -		
Db	1199	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	CC the European Bioinformatics Institute. There are no restrictions on its		
Qy	494	-----	493	CC	CC use by non-profit institutions as long as its content is in no way		
Db	1199	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	CC modified and this statement is not removed. Usage by and for commercial		
Qy	494	-----	493	CC	CC entities requires a license agreement (See http://www.isb-sib.ch/announce/		
Db	1199	SEGYARKQRNPKNQWSYVSKLYNEDPEVLDMDERLASRFFGKVRGRNYYEIPKKGYLL	1197	CC	or send an email to licensing@isb-sib.ch .		

Qy	997	RFGTIIISYIVLRSKGKISDRVILSSEYDPKHKYKDPYYTENQVLPAVRLIEAFGYRK			
Db	1618	RPGTVISYIVLKGSGRIGDRAIIPFDEDPKHKYDAEYYIENQVLPAVRLIEAFGYRK			
Qy	757	DLKQYQSSKQVGLDANLK 773			
Db	1678	DLRQKTKQVGLGAWLK 1694			
RESULT	12				
DPO1_THEHY	ID	DPO1_THEHY	STANDARD;	PRT;	1668 AA.
Q9HH0;	AC	AC			
DT	16-OCT-2001	[Rel. 4.0, Created)			
DT	16-OCT-2001	(Rel. 4.0, Last sequence update)			
DT	28-FEB-2003	(Rel. 4.1, Last annotation update)			
DE	DNA polymerase (EC 2.7.7.7) [contains: Endonuclease PI-ThyII (EC 3.1.1.-.) (Thy pol-1 intein); Endonuclease PI-Thy (EC 3.1.-.-.) (Thy pol-2 intein)] (Fragment).				
DE	POL.				
GN	Thermococcus hydrothermalis.				
OS	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;				
OC	Thermococcus hydrothermalis.				
OX	NCBI_TAXID:46539;				
RN	[1]	SEQUENCE FROM N.A.			
RA	Querellou J.J.E., Cambon M.A., Lesongeur F.O., Barbier G.;				
RA	"The Thermococcales system and phylogeny based on the comparative use of 16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA Polymerase genes".				
RT	Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.				
[2]	CHARACTERIZATION OF PI-THYI.				
RP	MEDLINE=20512590; PubMed=11058140;				
RA	Saves J., Blaume H., Dietrich J.J., Masson J.-M.;				
RA	"The Thy pol-2 intein of Thermococcus hydrothermalis is an isoschizomer of PI-ThyI and PI-ThyII endonucleases".				
RT	Journal of Bacteriology 182:4391-4396 (2000).				
RL	Nucleic Acids Res. 28:4391-4396 (2000).				
-1-	FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE EXHIBITS 3'- TO 5' EXONUCLEASE ACTIVITY (BY SIMILARITY).				
CC	-1- FUNCTION: PI-THYI AND PI-THYII ARE ENDONUCLEASES. PI-THYI CLEAVES THE INTEINLESS SEQUENCE OF THE THY DNA POL GENE. IT' REQUIRES A 2'-				
CC	BP MINIMAL RECOGNITION SEQUENCE.				
CC	-1- CATALYTIC ACTIVITY : N deoxynucleoside triphosphate = N diphosphate + {DNA} (N).				
CC	-1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN).				
CC	-1- FOLLOWED BY PEPTIDE LIGATION (Potential).				
CC	-1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.				
CC	-1- SIMILARITY: IN THE INTEIN SECTION, BELONGS TO THE HOMING ENDONUCLEASE FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation, the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.ebi.ac.uk/announce or send an email to license@ebi.ac.uk).				
CC	-----				
DR	EMBL; AJ245819; CAC18555.1; -.				
DR	HSPB; P56689; ITGO.				
DR	REBASS; 4332; PI-ThyI.				
DR	InterPro; IPR006172; DNA_pol_B.				
DR	InterPro; IPR006134; DNA_Pol_B-dom.				
DR	InterPro; IPR006133; DNA_Pol_B_exo.				
DR	InterPro; IPR003586; Hedgehog_HintC.				
DR	InterPro; IPR003587; Hedgehog_HintN.				
DR	InterPro; IPR006141; INTEIN.				
DR	InterPro; IPR006142; INTEIN.				
DR	InterPro; IPR004042; INTEIN_endonuc.				
DR	InterPro; IPR00578; Pol2.				

Db	DPOL_METVO	STANDARD;	PRT;	824 AA.
Qy	494	-----	493	
Db	718	VGTNGFRMRPFVIEEDPAKLLGYYSEGNARKWNQNGWSYTKLYNENQRVLDMD	777	
Qy	494	-----	493	
Db	778	SLAERFFGRVKRGINYIEIPKQAYIIFBNLCTLAENKRVPEAIFTSPESVWAPIEGY	837	
Qy	494	-----	493	
Db	838	FIGGDVHPSKRVRVLSTKSSELLYNGLVILLNLSGVSAAKIRDGTYRVVNEELPFTDY	897	
Qy	494	-----	493	
Db	898	RKKXNAYYSHVIPKEILEBTFGKYFORSVSYEKFRELVKSEKLDGEKAKRIEWLNGDV	957	
Qy	494	-----	516	NSYIGMGMGPKARWNSZCQESV
Db	958	LDKYLEVKKRKPYEGYYDLSVEEDENFLAGFLYIYAHNSYYGYGVARARYCKECMESV	101	
Qy	517	TAWGHYIENTIKE1BEKEFKFKVYAD-----	543	
Db	1018	TAWGDYIETTHE1BERGPKVYADSYTGETE1L1KRGKVEFVAIBELFQRDYRIG	1077	
Qy	544	-----	543	
Db	1078	EKEYCYLEGVEALLDNRGRBLWVKSVPYTMHRHTNKRYRVWPTNSWYLDVTDHSLLGY	113	
Qy	544	-----	543	
Db	1138	MNTSKVTPGKBLKERLVEVTPGEVSKSLITPNRAIAHGIRVNP1AVRLWBLIGLIVG	119	
Qy	544	-----	543	
Db	1198	DGMWGGOSNWAKEYANGSLGLDKBEEIEKILKPLRNTG1ISNYDKSKRGDYSILSRKMLA	125	
Qy	544	-----	543	
Db	1258	RFMVRYFKDESGSKRIPWMENLPREYIEAFRLGLFSADGTVSSLRKGYPEVRLTSYNP	131	
Qy	544	-----	543	
Db	1318	SSSVRKLIWLVGSNSMVFETVNPNRYLGKESGTHSVHYRIKDKTRFAERIGFLDLRKATK	137	
Qy	544	-----	547	--TDGF
Db	1378	LSENLGHTSKRKAKEYDFLVYPKVVEIAYDGYVDIEVEGTHRFANGLWHTNDGF	143	
Qy	548	YATIPGKEPKETIKKAKAREFLYKINSKLPGGLELEYEGFYLRGFVAKGRYADEGRT	607	
Db	1438	FAITPGDAETVKKAKFLKXINAKLPGGLELEYEGFYTRGFITKKKAVIADEGKLT	149	
Qy	608	TGCLEVYRDWSSEIAKETQAKWLEATLKDSEKRAVEIYVQDVEBIAKYQVPLKLYTHE	667	
Db	1498	TRELEIVRDRDWSSEIAKETQARLEA1RHDGEVEA1RVDVTBLSK1EVPPKLVH	155	
Qy	668	QITKDLSEYKAIGPHVIAKRLAAGK1KVRGPT1ISYV1LRSKGSKISDRVIL1SEYDPRK	727	
Qy	1558	QITRELDKYKATCPHVAIAKRLAARGKIRGSTVISYV1RKGSGRIGDRAIPFBDPFTK	161	
Qy	728	HRCDPDYIENQVLPAVRLAEGYRKEDLKYQSSKQVGDWALK	773	
Db	1618	HRDAEYIENQVLPAVRLKAFGKKEELRYQTRQVGLGAWLK	1663	

Qy	370	APNKDPDEEYRRRRLRTTYGGYKEPEGLWENITYLDFRCLYPSIIVTAVNSPDTL--E	427
Db	413	SPNPSSSSYRERAKFSTEGGYTREPLKGIOEDIVSDFMSLYPSIILISHNISPEVIE	472
Qy	428	REGCKQYDVAPIVYKFCKDFFPQIPISTLGELEMTPQBIKKKOMATDPIK-----K	480
Db	473	EKEBNMELL-----GILPKTNEELSSRRGHKMLK--DKIQKNEFDBEYS	516
Qy	481	MJDYQRQAVKLHANSYIGMYPKARWYSKECAESTAWGRHYIENITKETBKFGFKV	540
Db	517	RLHEEQOKSIKVULANSHYCYGLAFMARNYSDKCAEMVTGLGRKYIOETIEKAE-FERKV	575
Qy	541	YADDGFTATPGKPEPTEIKK-----AKEFLKTYINSKLPGLL	578
Db	576	YADDGFTAKWMDYKLQKGKKEENDKSDKLNSLPKLSKEELIILTAKPLGINEBPEGM	635
Qy	579	ELEEGFYURGFVAKGRYAVIDBEGRITTRGLEVVRDWSIAKETQAKYLEAIIKEDS	638
Db	636	ELEEGHFRGLFTKCKYALIEDGHLIVVKGLEVVRDWSIAKQVATRALEEDGD	695
Qy	639	VERKAEIYDVEEIAKYQVPLKLVTHEQITKDLSYEKAIGPHVIAJAKRAGSKVRP	698
Db	696	VNLAKKIKKNTIDNLUKGNIDRNLLIHTQLKNIEEVKSTAPIEYAKKIKQRGDSVR	755
Qy	699	GTISIYTIVRGSGCISDVEILISSEYDPCSKHCKYDPPDYYTENQVLPAVRILBEAFGTRKEDL	758
Db	756	GDVSYIIVKGSSRISERAELL-EY--AGDYDINYTIDNQVLPVIRIMESLG-SEDEL	811
Qy	759	KYQSKQVYGLDANL	772
Db	812	YKNCQKQVYGLDANL	821

RESULT 14					
DPOL_ARCFU	DPOL_ARCFU	STANDARD	PRT;	PRT;	781 AA.
AC	O29753;				
DT	15-TUL-1998	(Rel.	36	Created)	
DT	15-TUL-1998	(Rel.	36	Last sequence update)	
DT	28-BEL-2003	(Rel.	4	Last annotation update)	
		DNA polymerase	(EC 2.7.7.7).		
POL	OR	POLB OR	AF0497.		
OS		Archaeoglobus	fulgidus.		
OC		Archaea	Euryarchaeota	Archaeoglobi	Archaeoglobales;
OC					Archaeoglobus.

-1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
EMBL: AE001070; AAB90741.1; -;
DR PIR: A69312; A63312.
DR HSSP ; P566689; 1230.
DR TIGR ; AF0497; -;
DR Interzo; IPR000172; DNA_Pol_B.
DR Interzo; IPR000134; DNA_Pol_B_dom.
DR Interzo; IPR006133; DNA_Pol_B_exo.
DR Interzo; IPR00578; PolB.
DR Pfam; PF001336; DNA_Pol_B; 1.
DR Pfam; PF03104; DNA_Pol_B_exo; 1.
DR PRINTS; PRO0106; DNAPOLB.
DR SMART; SM00486; POLB; 1.
DR TIGRFRMS ; TIGR0592; polB; 1.
DR PROSITE; PS00116; DNA_Polymerase_B; 1.
DR Transferase; DNA-directed DNA_Polymerase; DNA replication;
CW DNA binding; Complete proteome.
SEQUENCE 781 AA; 89850 MW; 761C7B18FCC61B8B CRC64;
Query Match 32.3%; Score 1301; DB 1; Length 781;
Best Local Similarity 37.4%; Pred. No. 1..6..63; Gaps 21;
Matches 305; Conservative 155; Mismatches 266; Indels 90; Gaps 21;
2 1FDTDYLTQDGKPIIIRTFKRNENGEFKI1LDLPHFQPYIYALLKQDSDAIDEIKAIKGRHCK 61
8 LIDADIEFTIGKAVVRLWCKDDQGIFPAYDNDPFPYFVIGDE--DILKRNATSTRRE 64
62 IYRVVDAVKCKKFLGRDVEWKLIFPHQDYPALRQKIREHPAVIDIYEDYDPPFALKRYL 121
65 VIKLKSLSEPKAQLKTLKGLGEVEGYIVYAHQHPVQLRILYLSQPS--DVERADLPFAYRYL 121
122 IDKGIL----1PMEGDS-----ELKLMKAFDIEPTFYHEG-DEFG 153
122 IDKDLANDMDGIAEGEQQG3TYSKIEKVERIPRMFPELKLWVFCBEMLSFGMPPE 181
154 KGEIIMISYADEEEARYWITKWNIDLPVDDVSNEREMIKRKFQIVREKDPDVITYGDN 213
182 KDPITIVSVKTNDDDEI-----LTGDERKLTISDFYKLKISYDPPDLYVGNQDA 230
214 FDLPLVLRKRAEGLGVTLILLGRDKREHPPKIHNGDSSAVEKIRHPLFPVVRITMLP 273
231 FDWPYLRKRAEWNIPDVGRDGSN--VVPRGG--RPKITGRNLNDYDIAMRISDIK 284
274 TYTLBAYYEAVIGKTK-SKLGAEETAAIWETESEMKLQAYSMEDARATYELGKEFFPDM 332
285 IKKLENAAEFLGTTKIAEADKDIYIWSRGEK -EVLYNAYRQDA NTYLARELLLMH 343
333 AELAKLIGQSYWDSRSSTGNIWENYLLRVAYERNLALPNKEDEEYRRRLRTTYGGY 392
344 YELSNOMRLPYDVTWGRKQYDWTMSEARKNGEALPNPPEHAE-----SYBAEVFW 396
393 KEPERGLWENITYLDRFLCLYISLIVTHVNSDPTLREGCRN -YDAPIVGKFCFKDPG 450
397 LEPERGSHENTACLDPSAMYSIMIARNISPDTY--GCRDCCYE-APEVGHKFRKSPDGD 452
451 FIPSIIGLITMROEIKKMK-ATIDPIEKMKDLYDRAVFLKHANSYYGMYGPKARYTS 509
453 FPKRIRMLIIEKREKVELNLSPSESEYKLDIKQTLYLINTSFGYGMGNLARYC 512
510 KECAESVTAANGHYTEMTIKEIIEKCFKLYADTOGFYATPGKEKETIKKAKEFKLY 569
513 HPCAETATWAGRHFIRTSAK-TAESNGFKVQYDGDTSIFVTKAGMTEDVDR ---LIDK 567
570 INSKTPGULLEBEYEGCFYLRGFVAKFKEVAYAIDBEGRTTRTCLEVPRDOWSEIAKETTQAKY 629
58 LHEELD--IQEDEVDEYSAITFVEKRYAGLTDGRJLAVKGLEVRDWDCELAKKVQREV 625
630 LEAIIKEDSVKAETVYDQVVEIAKQVPLEKLVHETQITDSEKAIGPHVIAKRL 689
626 IEVILKEKNPKEKALSVKDVYTRIKGKVSUENVYKGTTKPKSYESEMAHVKALKL 685

